JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS **21**, 304-343 (1968)

On Stochastic Programming
I. Static Linear Programming under Risk

M. A. H. DEMPSTER[1]

Nuffield College, University of Oxford, Oxford, England

Submitted by Richard Bellman

1. INTRODUCTION

The theory of mathematical programming deals with the problem of maximizing or minimizing a function subject to constraints on the variables involved. In applications, however, these variables are often not the only ones to be considered, for certain parameters of the constraints or of the function to be optimized may be assumed to vary randomly. Such a situation gives rise to the concept of stochastic programming—a concept closely allied to parametric programming; i.e., the investigation of the behavior of the optimal value of a program as certain parameters are changed. In practice the distribution of the random variables involved in a stochastic program is almost never known, but rather must be estimated from available data. However as a first approximation, or in abstract applications, it is often expedient to assume that the required distributions are completely known and given. Following the distinction made by economists, the first case may be said to be programming under *uncertainty*, while the second may be called programming under *risk*.

It is the purpose of this paper to develop a way of looking at stochastic programming problems which is natural in statistical decision theory, to relate this approach to the previous research on linear programming under risk (in which it is implicit), and to make a detailed investigation of one type of stochastic linear programming problem within this framework. In much of the paper, all parameters of the linear program considered are allowed to be jointly distributed random variables with an arbitrary covariance structure. Furthermore, although computational details will not be considered, computational feasibility has been kept in mind.

Even in the simplest cases of stochastic programming, the problem involved is not immediately clear, for one has a whole class of programs depending on which values of the random variables are realized. Thus temporal con-

[1] Now at Balliol College, Oxford.

siderations are introduced into the problem and the question arises as to whether one must, or should, optimize before or after observing the random variables. If *before*, then once the random variables are realized there will usually be some loss or cost should the "optimal" decision violate the constraints. This possibility must be balanced against the favorable effects of solving the program in the deterministic sense when choosing the values of the variables under one's control. Since the value of the objective functional resulting from a given decision will be *a priori* a random variable, the nature of such favorable effects must be embodied in criteria for deciding among possible decisions specified in terms of this random value. On the other hand, if *after* the random variables are realized, one solves the resulting ordinary program, no prior decision or *stochastic optimality criterion* is involved. Instead, one is interested (often theoretically, or for purposes of comparison) in the distribution of the value of the linear program solved by this procedure.

It seems reasonable to make explicit the idea of making a prior decision so as to balance off a "good" value of the optimality criterion with a small expected value of a suitable loss function whose argument is the violation vector of the constraints. This formulation will be termed the *discrepancy cost* approach since the loss function represents the cost of any discrepancy between the decision vector and the constraints after the random variables have been realized. The idea is contained in much of the literature on stochastic programming, but it seems to have been first proposed explicitly (although not in this setting) by Theil [38].

There have been three principal formulations of stochastic programming problems (proposed roughly concurrently) and considerable discussion about the applicability and appropriateness of each. The first, which may be called the *distributional problem*, is associated with the names of G. Tintner [40] and H. Theil [39]. The linear case has been more recently investigated by colleagues of Tintner, e.g., [34, 36]. The second type, designated *chance-constrained programming*, has been proposed [11] and extensively studied [8, 9, 10] by A. Charnes and W. W. Cooper. The third formulation, termed "linear programming under uncertainty" in the literature, will be referred to as *stochastic programming with linear compensation* in this paper. It is due, independently, to E. M. L. Beale [3] and G. B. Dantzig [14]. Subsequently, it has been elaborated by Dantzig and A. Madansky [15, 31, 32, 33] and several other authors, e.g. [13, 45]. General discussions of stochastic programming have recently been presented [12, 33], and relatively comprehensive bibliographies have been given in several of the papers cited, particularly those of Charnes and Cooper. With relatively few exceptions only problems arising from linear programs under risk have so far been treated. In these investigations the right-hand sides of the linear inequality constraints are usually assumed to be random. Depending on the formulation, the parameters of the linear functional to be optimized or of the constraint matrix are sometimes assumed to be random as well.

Although this study deals explicitly with the minimal expected value criterion for static linear programming under risk, the discrepancy cost approach to stochastic programming is clearly applicable more generally. In the first place, the consideration of similar maximization problems and other tractable optimality criteria is straightforward. Secondly, possible extensions of the present results to nonlinear programs will be indicated at appropriate points below. The remaining restrictions are conceptually more difficult to relax. The restriction to the *static* case, i.e., at most two decision periods (before and after the random variables are realized), may however be removed in order to consider stochastic optimization problems over multiple periods under risk. It should be mentioned that Charnes, Cooper, A. Ben-Israel, and G. L. Thompson have made an extensive investigation of chance-constrained formulations for such k-stage problems utilizing the concept of "dynamic decision rules," see [12]. Finally the discrepancy cost approach is suitable for an attack on stochastic programming under uncertainty as a problem of statistical decision theory in a necessarily dynamic setting. These extensions will be the topics of future papers in this series.

After the notational preliminaries of Section 2, Section 3 introduces the discrepancy cost approach to stochastic programming rigorously. It is shown that while Tintner's prior decision version of the distributional problem involves zero discrepancy costs, Theil's problem, chance-constrained programming, and programming with linear compensation may all be nontrivially embedded in the class of problems of this type. In Section 4, quadratic loss is considered. Section 5 is a diversion which deals with convex polytopic cones and the pseudo- (or generalized) inverse of a matrix. As well as some classical results which will be needed, it contains new relations between the two theories which may be of independent interest. In Section 6, these notions are applied to finding necessary and sufficient conditions for the existence of solution vectors for stochastic programming with linear compensation. It turns out that the properties of a matrix connected with the compensating decision is crucial to the nature of the problem. The two most important cases, which lead respectively to constrained linear, and piece-wise linear losses, are treated in Sections 7 and 8. Although both sections contain new results, the latter is partly a collation of previous research placed in a new setting.[1]

[1] Parts of [13] have been of great value in applying techniques using the pseudo-inverse to the problems of Sections 6, 7 and 8.

2. Preliminaries

The following notational conventions will be observed in this paper. The relative complement of one point set Γ with respect to another Λ will be denoted by $\Lambda \backslash \Gamma$. The range, domain and nullity of a function f will be denoted respectively by $R(f)$, $D(f)$ and $N(f) = \{x: f(x) = 0, x \in D(f)\}$. The restriction of the function to the point set $\Gamma \subset D(f)$ will be denoted by $f \mid \Gamma$, while the image of this set under f will be denoted by $f[\Gamma]$. Further, for a point set $\Lambda \subset R(f)$, $\{x: f(x) \in \Lambda\}$ will be denoted by $f^{-1}[\Lambda]$.

This study will be concerned with operations in finite-dimensional Euclidean vector spaces, E_n, over the real field R. Scalars will as usual be denoted by small Greek letters, vectors by small Roman letters and matrices by Roman capitals. Both the scalar and vector zero will be denoted by the same symbol, 0. The topological notion of interior of a point set Γ in E_n will be denoted by int Γ. The elements of E_n will be thought of as column vectors and linear transformations between E_p and E_q will not be distinguished notationally from their matrices with respect to the usual orthonormal bases, e.g. $\{e_i, i = 1, \ldots, p\}$. Conjugation or transposition will be denoted by the prime symbol so that, for example, $e_2' = (0\ 1\ 0 \ldots 0)$. Functionals in linear programs will be written using the dot symbol for the inner product in E_n. The identity transformation on E_n will be denoted by I_n. A projection of E_n onto a closed linear subspace is here an idempotent matrix, while an orthogonal projection Q is a symmetric, idempotent matrix. The closed subspaces $Q[E_n]$ and $(I_n - Q)[E_n]$ associated with the orthogonal projections Q and $I_n - Q$ are othogonal complements; i.e., their direct sum is E_n and $x \cdot y = 0$ for all $x \in Q[E_n]$ and $y \in (I_n - Q)[E_n]$. They will be called a *complementary pair*.

A subset K of E_n is called a convex cone if u, $v \in K$, $\alpha, \beta \geqslant 0$ imply $\alpha u + \beta v \in K$. It is easy to check that K satisfies the usual definition of a convex point set. A partial order π may be induced on E_n through an arbitrary convex cone by defining, for $x, y \in E_n$, $x \pi y$ if and only if (iff) $x - y \in K$. In particular, if one takes K to be the closed positive orthant P_n in E_n, this partial order is simply the coordinate-wise partial order $\geqslant$. As usual for $x, y \in E_n$, $x > y$ and $x \gneqq y$ mean respectively $x - y \in \operatorname{int} P_n$ and $x - y \in P_n \backslash \{0\}$.

Throughout this paper fixed but arbitrary probability spaces (Ω, Σ, P) will be used as conceptual aids. The probability P will always be a Lebesgue-Stieltjes measure determined by a known (multivariate) distribution function, and usually assumed discrete, or absolutely continuous with respect to the Lebesgue measure of appropriate dimensions on the Euclidean sample space Ω. The expectation and variance operators, $E(\cdot)$ and $V(\cdot)$, respectively, will be used in the sense of Lebesgue-Stieltjes integration, but the term *Existence* for integrals will be used in the sense of *finiteness*. The terms random variable and random vector will sometimes be abbreviated respectively as r.v. and r.**v**. Random vectors occuring in stochastic programs will be denoted by the bold-face equivalents of the usual notation for the correspon-ding deterministic program and, for example, $\mathbf{b} = b$ will be used to denote a realization of the r.v. b.

The standard terminology of mathematical programming will be used with the following exceptions. A program with a finite optimum will be referred to as *proper*, so that in the contrary case a program will be referred to as *improper* rather than inconsistent or unbounded. If a deterministic mathematical program has the same set of optimal vectors as another program of comparable dimensions involving random variables, it will be said to be a *decision-equivalent* program for the latter. If the optimal values of the two programs, agree as well, the former will be said to be a *certainty-equivalent* for the latter. A *value-equivalent* program could also be defined, but this will not be needed. Often the distinction between a certainty-equivalent and a reformulation of the original program is rather vague.

3. The Discrepancy Cost Approach to Stochastic Programming

A deterministic mathematical programming problem is an optimization problem of the following type,

$$\min_{x \geqslant 0} v(x) \qquad \text{s.t.} \qquad G(x) \leqslant 0, \tag{3.1}$$

where $v : E_n \to E_k$ and $G : E_n \to E_m$. Usually $k = 1$, and in order to ensure the existence of global optima, v is a convex functional on E_n and G is a convex mapping. The letters s.t., subject to, in (3.1) mean that the minimiza-tion is to be carried out only over those $x \geqslant 0$ satisfying the condition $G(x) \leqslant 0$. Often these constraints are in the form of equations rather than inequalities, but it is well known that either form can be expressed in terms of the other. A deterministic linear program may be written as a special case of (3.1),

$$\min_{x \geqslant 0} c \cdot x \qquad \text{s.t.} \qquad (A\ b)\begin{pmatrix} -x \\ 1 \end{pmatrix} \leqslant 0 \qquad \text{or} \qquad Ax \geqslant b, \tag{3.2}$$

where A is an $m \times n$ matrix.

A stochastic programming problem involves functions of random vectors. Indeed, given an appropriate probability space (Ω, Σ, P), the functions v and G may be thought of as $v : E_n \times \Omega \to R$ and $G : E_n \times \Omega \to E_m$, of course assumed measurable with respect to Σ for arbitrarily fixed values of the other arguments. Then (3.1) and (3.2) become

$$\min_{x \geqslant 0} \mathbf{v}(x) \qquad \text{s.t.} \qquad \mathbf{G}(x) \leqslant 0, \tag{3.3}$$

and

$$\min_{x \geqslant 0} \mathbf{c} \cdot x \qquad \text{s.t.} \qquad \mathbf{A}x \geqslant \mathbf{b}, \tag{3.4}$$

where $\mathbf{v}(x)$ is a r.v. and $\mathbf{G}(x)$ is a r.v., for fixed x, and $\mathbf{c}$ and $\mathbf{b}$ are r.v.'s while $\mathbf{A}$ is a random matrix. One might think of choosing a vector from the convex polytope

$$\pi = \{x : x \geqslant 0,\ A(\omega)\, x \geqslant b(\omega),\ \text{all } \omega \in \Omega\}$$

to yield $\min_{x \in \pi} \sup_{\omega \in \Omega} c(\omega) \cdot x$. This is what Madansky has termed the "fat formulation" of stochastic linear programming [31, 33]. Often however, either π is vacuous, or the supremum does not exist.

The distributional problem for the program (3.4) may be formulated explicitly as follows. Suppose one "waits to see," i.e., suppose one waits until the triple $[\mathbf{c}, \mathbf{A}, \mathbf{b}] = [c, A, b]$ is realized, and then finds by ordinary linear programming the optimal solution $x^0[c, A, b]$. Then, beforehand, what is the distribution of $v(x^0[\mathbf{c}, \mathbf{A}, \mathbf{b}]) = c \cdot x^0[\mathbf{c}, \mathbf{A}, \mathbf{b}]$, its expected value, $Ev(x^0[\mathbf{c}, \mathbf{A}, \mathbf{b}])$, its variance, $V(v(x^0[\mathbf{c}, \mathbf{A}, \mathbf{b}]))$, etc.?

Tintner and his colleagues have supplied approximate answers to these questions and have been concerned with fitting data from certain resource allocation problems of agricultural economics in which the parameters of the objective function have the highest dispersion. [34, 36, 40]. Moreover, Tintner has proposed a decision problem in which one is required to find beforehand a certain $m \times n$ matrix U with entries the proportions of each of the m resources $\mathbf{b}$ allocated to each of the n activities whose matrix of resource input coefficients is $\mathbf{A}$ [41]. An appropriate maximization criterion is specified in terms of the vector of activity levels $x_U(\mathbf{A}, \mathbf{b})$ and their prices $\mathbf{c}$. Although the ordinary version of the problem is a linear program (see e.g. [27]), in the stochastic version U rather than x is the decision, and $x = x_U(A, b) \geqslant 0$ is the vector of activity levels not known with certainty until $[\mathbf{A}, \mathbf{b}] = [A, b]$ is realized. Thus there is no problem of constraint violation.

In the program (3.4) let A be fixed and either $\mathbf{c}$ or $\mathbf{b}$ be a r.v., but not both. Then the resulting distributional problem may be solved by applying the results of E. Simons on parametric programming [37]. Indeed, suppose that only $\mathbf{c}$ is random in the program (3.4).[2] Then the constraint set $\{x : x \geqslant 0, Ax \geqslant b\}$ is a closed (not necessarily bounded) convex polytope in E_n. It is well known that there are at most $\binom{m+n}{n}$ elements in the set Σ of

[2] This problem may be generalized as follows. Let Γ be a convex set in E_n and let X be a r.v. with known distribution taking values in E_n. Consider the r.v. $\varphi(X) = \sup\{X'z : z \in \Gamma\}$. What is the distribution of $\varphi(X)$ and that of the optimal $z(X)$, if it exists? Answers to these questions may be obtained for Γ an n-cube or ellipsoid quite simply. Another special case of this problem concerns Bayes risk in the statistical design of experiments.

extreme points of this polytope, say $\{x^k: k = 1,..., K\} \subset P_n$. Under the usual nondegeneracy hypothesis, these correspond to the basic feasible solutions of the simplex algorithm. In general the elements of Σ may be explicitly calculated for given A and b by simplex techniques. Simons has shown that to each x^k there corresponds a closed convex polytope $\Lambda(x^k) = \{c : c \text{ for which (3.2) is proper}\}$ in E_n. Hence an immediate necessary condition for $Ev(x^0[\mathbf{c}])$ to exist in this case is that

$$P\left\{\mathbf{c} \in \bigcup_{x^k \in \Sigma} \Lambda(x^k)\right\} = 1. \tag{3.5}$$

In principle, the distribution functions of the K r.v.'s $\mathbf{c} \cdot x^k$ can be calculated from the joint distribution function of the r.v. $\mathbf{c}$, and thus the distribution of $v(x^0[\mathbf{c}]) = \max_k\{\mathbf{c} \cdot x^k: k = 1,..., K\}$ determined. The proof of the following lemma (needed in Section 6) follows easily from the observation that for Lebesgue-Stieltjes measures integrability is equivalent to absolute integrability (since one is concerned with linear functions of the coordinates of $\mathbf{c}$).

LEMMA. 3.1. *For the program* (3.4) *with only the vector c random*, $Ev(x^0[\mathbf{c}])$ *exists iff* (3.5) *holds and* $E\mathbf{c}$ *exists.*

In the general distributional problem for (3.4), notice that except in trivial cases, $Ev(x^0[\mathbf{c}, \mathbf{A}, \mathbf{b}]) \neq v(x^0[E\mathbf{c}, E\mathbf{A}, E\mathbf{b}])$, the value of the deterministic program in which the random variables are replaced by their expected values. (This procedure is termed the "expected value" approach.) For the case of (3.3) which involves a deterministic quadratic functional v and linear equality constraints with random right-hand side $\mathbf{b}$; however, Theil has shown that the expected value procedure yields a decision-equivalent ordinary quadratic program which may be used to determine a decision $x \geqslant 0$ optimal with respect to minimizing the expected value of v before $\mathbf{b}$ is realized [39]. In applications the r.v. $\mathbf{b}$ represents the disturbances in an identified econometric model of an economy. It may be seen that this problem is a special case of the discrepancy cost program with quadratic loss treated in Section 4.

Therefore let us turn to stochastic programming problems more interesting from a decision-theoretic point of view and formulate explicitly the discrepancy cost approach with expected value optimality criterion. Formally, one replaces (3.3) by a minimization problem of the type

$$\min_{x \geqslant 0} E[\mathbf{v}(x) + L(\mathbf{G}(x))], \tag{3.6}$$

where $L : E_m \rightarrow R$ is the loss function representing the cost for any violation of the constraints by the decision x after the random variables are realized. Thus when the loss function is explicitly given, a discrepancy cost problem is a deterministic program. When v, G, and L are convex, (3.6) is convex.

Sometimes this program has only non-negativity constraints, but often the decision x is subject to additional nonstochastic constraints which theoretically, *but not computationally*, may usually be ignored. Such constraints will be introduced explicitly in Section 8 below. Otherwise, the reader may always supply appropriate constraints without being forced into anything but a straightforward modification of the argument. The linear version of (3.6) is

$$\min_{x \geqslant 0} E[\mathbf{c} \cdot x + L(\mathbf{A}x, \mathbf{b})], \tag{3.7}$$

where $L : E_m \times E_m \rightarrow R$. Often of course the loss function is simply a mapping from $E_m \rightarrow R$ of the form $L(b - Ax)$. In any case, the object is to make a decision x *before* the random variables are realized so as to minimize the expected losses.

It will now be shown that the second principle formulation of stochastic linear programming may be placed in this setting. This approach is called chance-constrained programming since it concerns constraints of the form

$$P_i\{a_i \cdot x \geqslant \boldsymbol{\beta}_i\} \geqslant \alpha_i\,, \qquad i = 1, \ldots, m, \tag{3.8}$$

or, formally, $P\{Ax \geqslant \mathbf{b}\} \geqslant \boldsymbol{\alpha}$, where A', a constant matrix, has columns $(a_1\, a_2 \ldots a_m)$, and the probabilities are taken with respect to the marginal distributions F_i of the co-ordinates $\boldsymbol{\beta}_i$ of the r.v. $\mathbf{b}$. Typically, $\alpha_i \geqslant \frac{1}{2}$ for $i = 1, \ldots, m$. Three main types of stochastic criterion functionals are minimized, so that three classes of programs involving the r.v. $\mathbf{c}$ result:

the E-type

$$\min_{x \geqslant 0} E\mathbf{c} \cdot x \qquad \text{subject to (3.8)}, \tag{3.9}$$

the P-type

$$\min_{x \geqslant 0} P\{\mathbf{c} \cdot x \geqslant \gamma\} \qquad \text{subject to (3.8)}, \quad \gamma \text{ fixed}, \tag{3.10}$$

and the V-type (H. Markovitz)

$$\min_{x \geqslant 0} E(\mathbf{c} \cdot x - \gamma)^2 \qquad \text{subject to (3.8)}, \quad \gamma \text{ fixed}, \tag{3.11}$$

although many others might be considered. Notice that if c is a fixed vector, the expectations and probabilities in (3.9), (3.10), and (3.11) are not required, since an x which satisfies (3.8) is not a random vector. In either case, the E-type chance-constrained program is already a deterministic (if somewhat complicated) program. Indeed, by writing (3.8) as

$$F_i(a_i \cdot x) \geqslant \alpha_i \qquad \text{or} \qquad a_i \cdot x \geqslant F_i^{-1}(\alpha_i), \qquad i = 1, \ldots, m, \tag{3.12}$$

(and defining for any distribution function F, $F^{-1}(\alpha) = \sup_\xi\{\xi : F(\xi) < \alpha\}$,) (3.9) is seen to be equivalent to the ordinary linear programme

$$\min_{x \geqslant 0} (E\mathbf{c}) \cdot x \qquad \text{s.t.} \qquad Ax \geqslant f, \qquad \text{where} \quad \varphi_i = F_i^{-1}(\alpha_i).$$

When A has a multivariate normal distribution and b is a fixed vector, a similar procedure yields quadratic constraints involving means, covariances and the standard normal distribution function Φ, see [8]. If c is a multivariate normal random vector, i.e., $c \sim N(m,\Sigma)$, the P-type program (3.10) may be reduced to the certainty-equivalent nonlinear program

$$\min_{x \geqslant 0} \left[1 - \Phi\left(\frac{\gamma - m \cdot x}{\sqrt{x'\Sigma x}}\right)\right] \qquad \text{s.t.} \qquad Ax \geqslant f. \tag{3.13}$$

While the not necessarily convex program (3.13), and more generally (3.9), may be treated by the calculus of variations, under similar normality assumptions S. Kataoka has reduced an alternate formulation of the P-type,

$$\min \gamma \qquad \text{s.t.} \qquad P\{\mathbf{c} \cdot x > \gamma\} = \beta \qquad \text{and (3.8)}, \quad \beta \text{ fixed}, \tag{3.14}$$

to a certainty-equivalent first-order homogeneous program [26]. It follows quite easily from the Kuhn-Tucker theory that the dual of this program is a linear program whose solution must satisfy a length constraint [18].

Charnes and Cooper's k-stage extensions of chance-constrained programs are problems in which in each period a decision x^t must be made concerning a stochastic linear program which involves past decisions and realizations of the $\mathbf{b}^t$. The decision rule may be required to come from a class of rules of a specified mathematical form (e.g. linear) so that it is not necessarily optimal in the class of all decision rules.

All these formulations have an obvious analogy in the Neyman-Pearson theory of hypothesis-testing in mathematical statistics. Indeed, under certain circumstances the introduction of a randomized decision may improve the expected value of (3.9). For example, consider the following special case of (3.9). Let $c > 0$ and $A = a > 0$ be real numbers and let $\mathbf{b}$ be a random variable which takes two values $b_1 > b_2 > 0$ with probabilities $p > 0$ and $1 - p$. Further suppose $0 < \alpha < 1$ and $1 - p < \alpha < p$. Then (3.9) becomes the simple program

$$\min_{x \geqslant 0} cx \qquad \text{s.t.} \qquad P\{ax \geqslant \mathbf{b}\} \geqslant \alpha, \tag{3.15}$$

and the single constraint $ax \geqslant F_b^{-1}(\alpha) = b_1$ is binding so that the optimal solution is $x_1^0 = b_1/a$ and the value of (3.15) is $v(x_1^0) = cb_1/a$. However, $v(x_1^0) > 0$, so that the introduction of a randomization device with probabilities α and $1 - \alpha$ to choose between the decisions $x_1{}^0$ and $x = 0$ will give rise to a randomized decision $X_1{}^0$ which satisfies the constraint of (3.15) and yields the expected value

$$cEX_1{}^0 = c(\alpha x_1 + (1 - \alpha)0) = \alpha c x_1{}^0 < c x_1{}^0.$$

Since $v(x_1{}^0) > v(x_2)$, where $x_2 = b_2/a < x_1{}^0$, an alternate feasible randomized decision $X_2{}^0$ is to choose $x_1{}^0$ with probability β and x_2 with probability $1 - \beta$, where $1 - p + \beta p = \alpha$. The expected value of (3.15) with this decision is

$$cEX_2{}^0 = c(\beta x_1{}^0 + (1 - \beta)x_2) \leqslant c x_1{}^0.$$

Which of these decisions would be better depends of course on the actual values of the parameters. In applications, the introduction of a randomized decision may be undesirable.

Perhaps more serious, since the resulting optimal decision rules are occasionally somewhat unusual, is the discontinuous nature of the loss functions implicit in chance-constrained programming. It is easily seen that loss is, in effect, zero for any $x \geqslant 0$ satisfying the constraint (3.8) and infinite for any other $x \geqslant 0$. Hence for such a loss function, a chance-constrained program reduces to a discrepancy cost program.

Before relating stochastic programming with linear compensation to the discrepancy cost approach, some natural loss functions will be introduced. Of course, many others are possible. Notice first that a linear loss function defined by $L(b - Ax) = f(b - Ax)$ will lead to an improper program

$$\min_{x \geqslant 0} E[\mathbf{c} \cdot x + f \cdot (\mathbf{b} - \mathbf{A}x) = \min_{x \geqslant 0} E(\mathbf{c} - \mathbf{A}'f) \cdot x + f \cdot E\mathbf{b},$$

unless either $E(\mathbf{c} - \mathbf{A}'f) \geqslant 0$, when the program is trivial, or else x is further constrained. It will be seen in Section 7 that, in certain instances, programming with linear compensation involves just such a constrained linear loss.

Next consider a loss function which is of the first order, i.e., the sum of a linear and a first order homogeneous loss function defined by

$$L(b - Ax) = f \cdot (b - Ax) + \| D(b - Ax) \|, \tag{3.16}$$

where D is an arbitrary $p \times m$ matrix and A is assumed to be fixed. That is, the program (3.7) is specialized to

$$\min_{x \geqslant 0} E(\mathbf{c} \cdot x + f \cdot (\mathbf{b} - Ax) + \| D(\mathbf{b} - Ax) \|. \tag{3.17}$$

When $f = 0$, the loss is symmetric on either side of the constraints—usually an undesirable feature in applications. The following results concern the program (3.17). The reader is referred to [17] for their proofs, but it should be mentioned here that the methods depend on the linearity of the constraints and the functional to be minimized.

THEOREM 3.2. *If the expectation of* $\mathbf{c}$ *and the covariance matrix of* $\mathbf{b}$ *exist, then either of the following conditions are sufficient for the discrepancy cost program* (3.77) *to be proper:*[3]

(i) *there exists a vector* $u \in E_p$ *such that* $\| u \| \leqslant 1$ *and* $E\mathbf{c} - A'f \geqslant A'D'u$, or

(ii) *the* $n \times n$ *matrix* $A'D'DA - (E\mathbf{c} - A'f)(E\mathbf{c} - A'f)'$ *is nonnegative definite.*

Under the conditions of Theorem 3.2, the pseudo-inverse[4] A^+ of the matrix A may be used to obtain bounds for the expected value of (3.17). Generally speaking, the smaller the covariance matrix of $\mathbf{b}$, the better the approximation of the following theorem.

THEOREM 3.3. *If the conditions of Theorem* 3.2 *hold, then the value* v^0 *of the discrepancy cost program* (3.17) *satisfies*

$$| v^0 - E\mathbf{c} \cdot A^+E\mathbf{b} - f \cdot (I - AA^+)E\mathbf{b} \,| \leqslant E \,\| D(\mathbf{b} - AA^+E\mathbf{b}) \|, \quad (3.18)$$

the upper bound being attained for $x = A^+E\mathbf{b}$ *(not necessarily feasible). If* $E\mathbf{b} \in R(A)$, *then* (3.18) *becomes simply*

$$| v^0 - E\mathbf{c} \cdot A^+E\mathbf{b} \,| \leqslant E \,\| D(\mathbf{b} - E\mathbf{b}) \|. \quad (3.19)$$

From Theorem 3.2, we see that one must again consider the nature of $Ec - A'f$ to determine whether or not the first-order version of (3.8) is proper (since the loss is of the same order as the function to be minimized). This and the standard two-term Taylor series approximation for nonlinear functions suggest the consideration of quadratic loss. Indeed, the following version of (3.7) will be investigated in Section 4,

$$\min_{x \geqslant 0} E(\mathbf{c} \cdot x + f \cdot (\mathbf{b} - \mathbf{A}x) + \| D(\mathbf{b} - \mathbf{A}x) \|^2), \quad (3.20)$$

with D as in (3.16).

[3] The condition (i) stated previously was slightly stronger in that it ruled out equality. However, using the original version of the Kuhn-Tucker theorem instead of the Slater-Uzawa form [44], the stronger result follows easily. It would appear that weaker necessary conditions would be considerably more complicated, and in practice not very useful, cf. [17].

[4] In Section 5 references to the literature on pseudo-inverses are cited.

The loss functions in the programs (3.17) and (3.20) assume that the linear programming problem giving rise to the stochastic problem has equality constraints (although with different costs of violation on either side of the constraints). In many applications this is actually the case, and it would appear that nonlinear loss functions are suited to this type of problem. On the other hand, the case of inequality constraints is better handled by piece-wise linear loss functions, since loss is then usually zero on one side of the constraints. Such loss functions are defined by $L(b - Ax) = \min_r \{f_r(b - Ax)\}$; and in the case of a fixed constraint matrix, we shall see that such a loss is generated in certain instances of programming with linear compensation.

The discrepancy cost approach to this formulation may be conceived as follows. Consider a loss function which is itself the value of the deterministic linear program defined by

$$L(b - Ax) = \min_{y \geqslant 0} f \cdot y \qquad \text{s.t.} \qquad By = b - Ax. \tag{3.21}$$

Hence, after $[\mathbf{A}, \mathbf{b}] = [A, b]$ is realized, a compensating decision $y_x(A, b)$ contingent on the choice of x, and on A, b must be made, and (3.7) becomes

$$\min_{x \geqslant 0} E(\mathbf{c} \cdot x + \min_{y \geqslant 0} f \cdot y) \qquad \text{s.t.} \qquad \mathbf{A}x + By = \mathbf{b}, \tag{3.22}$$

where $\mathbf{A}$ is a random $m \times n_1$ matrix, B is an $m \times n_2$ matrix, $\mathbf{c}$ is a random n_1 vector and f is an n_2 vector. The n_1 and n_2 vectors x and y are thus to be chosen in two stages to solve (3.22). Depending on the nature of the matrix B, i.e., the form of a certain polytopic cone that it generates, (3.21) is either a constrained linear, or a piece-wise linear loss function or a combination of both. The constrained linear case will be treated below in Section 7 and the other cases discussed in Section 8.

4. Quadratic Loss

In this section we study briefly the discrepancy cost program (3.20) with loss function defined by

$$\begin{aligned} L(b - Ax) &= f \cdot (b - Ax) + \| D(b - Ax) \|^2 \\ &= f \cdot (b - Ax) + (b - Ax)' H(b - Ax), \end{aligned} \tag{4.1}$$

where $H = D'D$, and is therefore a symmetric, non-negative definite matrix. The programming problem involved, letting $d = E\mathbf{c} - E\mathbf{A}'f$, is

$$\min_{x \geqslant 0} Ew(x, \mathbf{A}, \mathbf{b}) = \min_{x \geqslant 0} (d \cdot x + E[(\mathbf{b} - \mathbf{A}x)' H(\mathbf{b} - \mathbf{A}x)]. \tag{4.2}$$

The term $f \cdot E\mathbf{b}$ must be added to regain (3.20). (For the moment all expectations required are assumed to exist.) Now for fixed x, A and b, $w(x, A, b)$ may be expanded as $w(x, A, b) = (d - 2A'Hb)\, x + x'A'HAx + b'Hb$. Hence the programming problem involved in (4.2) is, letting

$$s = d - 2E(\mathbf{A}'H\mathbf{b}) = E\mathbf{c} - E\mathbf{A}'f - 2E(\mathbf{A}'H\mathbf{b}),$$

$S = E(\mathbf{A}'H\mathbf{A})$ and $\varphi(x) = Ew(x, \mathbf{A}, \mathbf{b}) - E(\mathbf{b}'H\mathbf{b})$, the ordinary quadratic program

$$\min_{x \geqslant 0} \varphi(x) = \min_{x \geqslant 0} (s \cdot x + x \cdot Sx), \tag{4.3}$$

which may be solved by the usual means for the optimal solution x^0 (see, e.g., [24]). Then x^0 is the solution of (3.20) whose value is

$$[E\mathbf{c} - E\mathbf{A}'f - 2E(\mathbf{A}'D'D\mathbf{b})] \cdot x^0 + E \| D\mathbf{A}x^0 \|^2 + E \| D\mathbf{b} \|^2 + f \cdot E\mathbf{b}. \tag{4.4}$$

Now since $H = D'D$ is non-negative definite, so is $A'HA$ for fixed A. For random $\mathbf{A}$, since $x'\mathbf{A}'H\mathbf{A}x$ is a convex function of the r.v. $\mathbf{A}x$, and using Jensen's inequality (see, e.g. [47], Section 4.14, pp. 67–70),

$$x'E(\mathbf{A}'H\mathbf{A})x = E(x'\mathbf{A}'H\mathbf{A}x) \geqslant x'E\mathbf{A}'HE\mathbf{A}x \geqslant 0, \tag{4.5}$$

for all $x \in E_n$. Hence $E(\mathbf{A}'H\mathbf{A})$ is non-negative definite, φ is a convex function, and the program (4.3), and thus the discrepancy cost program (3.20), is always proper.[5]

Notice that if the covariance matrix of $\mathbf{b}$ exists, its trace, hence $E \| \mathbf{b} \|^2$, and thus $E \| \mathbf{b} \|$, also exist. If $\| D \|$ is the Euclidean vector norm for D, this implies that $E \| D\mathbf{b} \| \leqslant \| D \| \, E \| \mathbf{b} \| < \infty$, and hence that the last two terms of (4.4) exist (see [25], Section 2.2, pp. 39-45). A similar assumption for $\mathbf{A}$ as a random vector in E_{mn} guarantees the existence of the terms that involve it. Indeed, for arbitrary $x \in E_n$, $E(\mathbf{A}'D'D\mathbf{b}) \cdot x \leqslant \| D \|^2 E(\| \mathbf{A} \| \, \| \mathbf{b} \|) \, \| x \|$. By the Schwartz inequality, $E(\| \mathbf{A} \| \, \| \mathbf{b} \|) \leqslant (E \| \mathbf{A} \|^2 E \| \mathbf{b} \|^2)^{1/2} < \infty$. Similarly, $E \| D\mathbf{A}x \|^2 \leqslant \| D \|^2 E \| \mathbf{A} \|^2 \| x \|^2 < \infty$ as required. Thus

THEOREM 4.1. *If the expectation of* $\mathbf{c}$ *and the covariance matrices of* $\mathbf{A}$ *and* $\mathbf{b}$ *exist, the discrepancy cost program* (3.20) *is proper, has a decision equivalent program* (4.3), *and its value is of the form* (4.4).

The relatively simple nature of (4.3) is a consequence of the linearity of the function to be minimized. If the program (3.20) had involved a nonlinear

[5] I am indebted to Professor M. M. Rao of the Carnegie Institute of Technology for this simplification of a previous argument.

function and linear constraints, more stringent conditions on the distribution of the r.v. **c** would have been required and a more complicated nonlinear program would have replaced (4.3).

5. Convex Polytopic Cones, Linear Inequalities and Pseudo-Inverses

In order to facilitate the study of stochastic programming with linear compensation, this section deals with the theory of convex polytopic cones[6] and the related inequalities which generate them. Several fundamental results of this theory, including the Minkowski-Weyl and Minkowski-Farkas theorems, are stated and it is shown that the mapping properties of the pseudo-inverse of a matrix [5–7, 17] are related to special cases of these classical theorems.

The definition of a convex polytopic cone will be introduced through a simple translation lemma about the image of the positive orthant P_n of E_n under an $m \times n$ matrix G. Unless stated otherwise in the following, the matrix G will be arbitrary. The lemma will be needed in Section 6.

For the $m \times n$ matrix G let $R_+(G)$ denote $G[P_n]$, and $R_-(G)$ denote $G[-P_n]$. Since $R_+(G) = \{v : v = Gu,\ u \geqslant 0\}$, it is easy to see that $R_+(G) = -R_-(G)$, i.e., that

$$v \in R_+(G) \quad \text{iff} \quad -v \in R_-(G). \tag{5.1}$$

Lemma 5.1. *Let the vector* $w \in E_m$. *Then* $R_+(G) + w \subset R_+(G)$ *iff* $w \in R_+(G)$. *Similarly,* $R_-(G) + w \subset R_-(G)$ *iff* $w \in R_-(G)$.

Corollary 5.2. *Both* $R_+(G)$ *and* $R_-(G)$ *are convex cones.*

The proof of the lemma, its corollary, and Lemma 5.3 below are straightforward and will be ommitted (see [17]).

Notice that $R_+(G)$ and $R_-(G)$ may be thought of, respectively, as the sets generated by all positive, and all negative, combinations of the column vectors of the matrix G and hence are polytopes in E_m. In the following definition either $R_+(G)$ or $R_-(G)$ could have been used, but the choice of $R_+(G)$ is somewhat more convenient for later purposes. A set Γ will be said to be a *convex polytopic cone* in E_m if there exists an $m \times n$ matrix G, for some n, such that $\Gamma = R_+(G)$.

[6] In the literature these cones are usually called 'polyhedral,' but it has been pointed out (see [22], p. xii) that the suffix 'tope' is the proper one for E_m. After they are defined, these sets will simply be called cones, since they are the only kind under consideration.

Most of the following terminology is standard [16, 21–23, 43]. Geometrically, $R_+(G)$ may be said to be the sum of n *half-lines* in E_m. The jth column g_j of the matrix G is a point of the jth half-line of this sum. The minimal number of such half-lines required to generate $R_+(G)$ is of course greater than or equal to the rank of G. Any set of half-lines which generate a cone in this fashion are said to *span* the cone, since the cone is simply their convex hull. A minimal set of r columns of the matrix G forming the $m \times r$ matrix H for which $R_+(H) = R_+(G)$ will be said to be a *frame* of the cone $R_+(G)$. Thus a frame is here a (not necessarily minimal) set of half-lines which span the cone. The *dimension* $d(G)$ of a cone $R_+(G)$ is the dimension of the smallest subspace $D(G)$ of E_m containing $R_+(G)$. Since $D(G) = R(G)$, this subspace may of course be generated by any maximal linearly independent set of columns of G. The *lineal space* $L(G)$ of $R_+(G)$ is the largest subspace of $E_m \subset R_+(G)$. The dimension $l(G)$ of $L(G)$ is said to be the *lineality* of $R_+(G)$. If $L(G) = R_+(G) = E_m$, $R_+(G)$ is said to be a *solid* cone. On the other hand, if $L(G) = \{0\}$ so that $l(G) = 0$, $R_+(G)$ is said to be a *pointed* cone.

Lemma 5.1 is a statement about the containment of the translate of a cone in the cone itself. For an $m \times n$ matrix G, the following lemma tells when a translate of the cone $R_+(G)$ is equal to $R_+(G)$.

LEMMA 5.3. *Let the vector* $w \in E_m$. *Then* $R_+(G) + w = R_+(G)$ *iff* $w \in L(G)$.

The next result has recently been shown to be fundamental to the theory of convex polytopic cones and linear inequalities in finite-dimensional inner-product spaces by A. Ben-Israel [4].

THEOREM 5.4. *Let* $\{M, N\}$ *be a complementary pair of subspaces in* E_n. *Then the following situations (perhaps after a relabelling of coordinates) are mutually exclusive*:

(i) $N \cap P_n = \{0\}$, *when* M *has a basis in int* P_n.

(ii) $N \cap P_n = R_+(e_1, ..., e_p)$, *when* M *has a basis in int* $R_+(e_{p+1}, ..., e_n)$.

(iii) N *has a basis in int* $R_+(e_1, ..., e_q)$, q *minimal, but*

$$N \cap P_n \neq R_+(e_1, ..., e_q), \qquad \textit{when} \quad M \cap P_n = R_+(e_{q+1}, ..., e_n).$$

(iv) $N \cap \text{int } P_n \neq \emptyset$, *when* $M \cap P_n = \{0\}$.

Note that it is possible that $M \cap P_n = R_+(e_{p+1}, ..., e_n)$ in condition (ii). When the subspace M intersects the positive orthant in a particular way under conditions (i), (ii), or (iii), one can prove a result concerning the decomposition of a vector $x \in P_n$ into components in M and N which is important for the study of stochastic programming with linear compensation and for the theory of pseudo-inverses.

THEOREM 5.5. *Let $\{M, N\}$ be a complementary pair of subspaces in E_n. Then all $x \in P_n$ may be uniquely decomposed as $x = x^1 + x^2$, $x^1 \in M \cap P_n$, $x^2 \in N$, if M has an orthogonal basis in P_n.*

PROOF: The case $M = E_n$ is trivial, so suppose $\dim(M) = m < n$ and let $\{y_1, \ldots, y_m\} \subset M \cap P_n$ be an orthogonal basis for M. Define the set $\{Q_1, \ldots, Q_m\}$ of mutually orthogonal projections such that

$$Q_i x = \alpha_i(x) y_i, \qquad -\infty < \alpha_i(x) < \infty, \qquad i = 1, \ldots, m.$$

Then the sum of any subset of this set is an orthogonal projection and, in particular, $Q = \sum_{i=1}^{m} Q_i$ and $I - Q$ are the projections associated with M and N, respectively.
For arbitrary $x \geqslant 0$,

$$x^1 = Qx = \sum_{i=1}^{m} \alpha_i y_i \qquad \text{and} \qquad x^2 = (I - Q)x,$$

give a unique decomposition. Since $y_i \geqslant 0$ for all i, it suffices to show that $\alpha_i \geqslant 0$ for all i. Hence consider $Q_i x = \alpha_i y_i$ for some $1 \leqslant i \leqslant m$ and assume $\alpha_i < 0$. Then x may be decomposed as

$$x = Q_i x + (I - Q_i)x \geqslant 0,$$

so that

$$(I - Q_i)x \geqslant -Q_i x = -\alpha_i y_i \geqslant 0.$$

But this contradicts the orthogonality of $Q_i x$ and $(I - Q_i) x$ and so $\alpha_i \geqslant 0$. The converse of Theorem 5.5 is obvious for $n = 1, 2, 3$. It might be conjectured to hold for all finite n.

We now turn to the relations between the pseudo-inverse and cone theory. Theorem 5.5 has an application to the projections associated with the (unique) $n \times m$ pseudo-inverse of an $m \times n$ matrix G. It is well known that G^+G and $I - G^+G$ are the orthogonal projections associated with the complementary pair $\{R(G'), N(G)\}$, while GG^+ and $I - GG^+$ are those associated with $\{R(G), N(G')\}$.

COROLLARY 5.6. *Let G be an $m \times n$ matrix. Then if there is an orthogonal basis for $R(G')$ in P_n, $G^+G[P_n] \subset P_n$. Similar statements apply to $N(G)$, $R(G)$ and $N(G')$.*

The conclusion of corollary 5.6 is trivially equivalent to the non-negativity of G^+G. For example, if

$$G = \begin{pmatrix} -1 & -2 & 2 \\ 2 & 4 & 0 \end{pmatrix},$$

then

$$G^+ = \frac{1}{10}\begin{bmatrix} 0 & 2 \\ 0 & 4 \\ 10 & 5 \end{bmatrix}$$

and

$$G^+G = \frac{1}{5}\begin{bmatrix} 1 & 2 & 0 \\ 2 & 4 & 0 \\ 0 & 0 & 5 \end{bmatrix}.$$

Here,

$$\left\{\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right\}$$

is an orthogonal basis for $R(G')$ in P_3. The non-negativity of G^+G always holds for matrices G of full column rank, since then $G^+ = (G'G)^{-1}G'$ and $G^+G = I$, so that $G^+G[P_n] = P_n$. The conjecture concerning the converse of Theorem 5.5, equivalently, the truth of the converse of Corollary 5.6, is supported by the fact that the non-negativity of G^+G does not even hold for (non-negative) matrices G whose rows are the frame of a pointed cone in P_n. For example, let

$$G = \begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & 2 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}.$$

Then

$$G^+ = \frac{1}{12}\begin{bmatrix} 3 & -3 & 6 \\ -3 & 3 & 6 \\ -1 & 5 & -2 \\ 5 & -1 & -2 \end{bmatrix}$$

and

$$G^+G = \frac{1}{4}\begin{bmatrix} 3 & 1 & -1 & 1 \\ 1 & 3 & 1 & -1 \\ -1 & 1 & 3 & 1 \\ 1 & -1 & 1 & 3 \end{bmatrix}.$$

When G^+G is non-negative, the defining property of the set of points in the cone $R_+(G)$ may be stated as, essentially, a set of n inequalities. First a general duality theorem will be proved.

THEOREM 5.7. *Let G be an $m \times n$ matrix and G^+ be its unique $n \times m$ pseudo-inverse. Then the following equality between sets in E_m holds.*

$$\{v : v = Gu, u \geqslant 0, u \in R(G')\} = \{v : G^+v \geqslant 0, v \in R(G)\}. \tag{5.2}$$

PROOF: Let the first set of (5.2) be Γ_1 and the second be Γ_2, and choose $v \in \Gamma_1$. Then $v = Gu$, $u \geqslant 0$, $u \in R(G')$, and hence $G^+v = G^+Gu = u \geqslant 0$, since $u \in R(G')$ and G^+G is the orthogonal projection onto $R(G')$ along $N(G)$. Therefore $v \in \Gamma_2$ and $\Gamma_1 \subset \Gamma_2$. Now let $w \in \Gamma_2$ be arbitrary. Then by the mapping properties of the pseudo-inverse, $G^+w \geqslant 0$ belongs to $R(G')$ and $w \in R(G)$. Since GG^+ is the orthogonal projection onto $R(G)$ along $N(G')$, $GG^+w = w$. Therefore $w \in \Gamma_1$, $\Gamma_2 \subset \Gamma_1$ and the proof is complete.

COROLLARY 5.8. *If G is an $m \times n$ matrix such that G^+G is non-negative, then*

$$R_+(G) = \{v : G^+v \geqslant 0, v \in R(G)\}. \tag{5.3}$$

PROOF: It follows from Theorem 5.7 that one always has

$$\{v : G^+v \geqslant 0, v \in R(G)\} \subset R_+(G). \tag{5.4}$$

On the other hand, by assumption, and Theorem 5.7,

$$\begin{aligned} R_+(G) &= \{v : v = Gu, G^+Gu \geqslant 0, u \in P_n\} \\ &\subset \{v : v = Gu, G^+Gu \geqslant 0, u \in E_n\} \\ &= \{v : G^+v \geqslant 0, v \in R(G)\}, \end{aligned}$$

using the mapping properties of the pseudo-inverse.
Similar results to Theorem 5.7 and its corollary may be obtained for more general vector spaces [19].

Corollary 5.8 is closely related to the classical Minkowski-Weyl theorem which says that every convex polytopic cone Γ, which is the convex hull of a finite number of half-lines, is the intersection of a finite number of half-spaces. The theorem may be stated in matrix form as follows (cf. [15] or [16]).

THEOREM 5.9. *Let G be an $m \times n$ matrix. Then there exists a matrix D such that the set Γ,*

$$\{v : Dv \geqslant 0\} = R_+(G), \tag{5.5}$$

and each column of D' is orthogonal to n-1 columns of G. That is, each column is a set of direction parameters of a hyper-plane which lies on the boundary of a half-space. Conversely, given a matrix D and set Γ of (5.5), there exists a matrix G such that $\Gamma = R_+(G)$.

Theorem 5.9 is usually proved by induction (see e.g. [23]). A.J. Goldman and A.W. Tucker have given a proof of the converse part based on a study of the face structure of the cones involved [28]. H. Uzawa has given a recursive procedure for calculating a frame of the cone when it is given in the form of the intersection of half-spaces, i.e., the set Γ of (5.5) is given [43]. Notice that the first part of Theorem 5.9 asserts the existence of a matrix D with the required properties. Its column dimension must be m but its row dimension is unspecified. The requirement of (5.3) that $v \in R(G)$ in Corollary 5.8 is necessary in order to exclude the vectors w of $N(G')$, the orthogonal complement of $R(G)$, which satisfy the equation $GG^+w = 0$. Therefore, after finding any solution w of the inequality $G^+w \geqslant 0$ of (5.3) (see [39] for methods), a vector $v \in R_+(G)$ may be obtained by taking $v = GG^+w$. Thus when G is such that G^+G is non-negative, taking

$$D = \begin{bmatrix} G^+ \\ I - GG^+ \\ GG^+ - I \end{bmatrix},$$

and using Corollary 5.8 yields a special case of Theorem 5.9. In particular, if G is of full column rank n, then $R(G') \cap P_n = P_n$. On the other hand, notice in the proof of the corollary that when $R(G') \cap P_n = \{0\}$, the containment (5.4) becomes trivial, viz. $\{0\} \subset R_+(G)$.

The Minkowski-Weyl theorem may be stated using the notion of the polar cone of a set. From this formulation, the second fundamental theorem of convex polytopic cones follows easily [21]. Known as the Minkowski-Farkas lemma, this theorem also follows from a corollary of Theorem 5.4 due to D. Gale and Tucker [4, 28]. The strong part of the Minkowski-Farkas lemma has several equivalent statements, two of which will be needed below.

Theorem 5.10. (1) *Let G be an $m \times n$ matrix and $v \in E_m$. Then $v \cdot w \geqslant 0$ for all solutions of the inequality $G'w \geqslant 0$, iff $v \in R_+(G)$.*

(2) *If G is an $m \times n$ matrix such that $R_+(G)$ is not solid, i.e., $E_m \backslash R_+(G) \neq \emptyset$, then there exists a hyperplane separating $u \in E_m \backslash R_+(G)$ and $R_+(G)$. That is, there exists a vector $w \in E_m$ such that $w \cdot v \geqslant 0$ for all $v \in R_+(G)$ and $w \cdot u < 0$.*

The second statement is particularly useful in analysis, and as a geometric form of the Hahn-Banach theorem is true for more general spaces.

Finally, we turn to a consideration of solid cones, or alternately, $m \times n$ matrices G such that $R_+(G) = E_m$. A set of n vectors $\{g_i\}$ in E_m will be said to be ***positively independent*** if no g_i lies in the convex hull of the others, i.e. if there does not exist a relation of the form

$$g_i = \alpha_1 g_1 + \cdots + \alpha_{i-1} g_{i-1} + \alpha_{i+1} g_{i+1} + \cdots + \alpha_n g_n , \tag{5.6}$$

with $\alpha_j \geqslant 0$, $j = 1,\ldots, i-1, i+1,\ldots, n$ (cf. [16, 29, 35]). Alternately, this can be looked on as

$$g_i \notin R_+(g_1, \ldots, g_{i-1}, g_{i+1}, \ldots, g_n), \qquad i = 1,\ldots, n. \tag{5.7}$$

Any frame of an arbitrary cone $R_+(G)$ is a set of positively independent vectors.

In stochastic programming with linear compensation, as will be seen in Section 8, one is interested in the frames of solid cones, i.e., sets of positively independent vectors which are the columns of a matrix G such that $R_+(G) = E_m$. Necessary conditions for the n columns of G to be such a set, called a *positive basis*, are given in the following theorem of C. Davis, applied to the case when G is of full row rank m [16]. These conditions are also sufficient when, in addition, the columns of G are positively independent.

THEOREM 5.11. *Let G be an $m \times n$ matrix. Then the following conditions are equivalent*:

(i) $L(G) = R_+(G) = R(G)$.

(ii) $N(G) \cap \operatorname{int} P_n \neq \emptyset$.

(iii) *Every vector $-g_i$ lies in the convex polytopic cone generated by the other $n - 1$ column vectors $\{g_i\}$ of G (cf. (5.7)).*

Davis has shown that under the conditions of the theorem the intrinsic geometry of G is revealed by a study of the face structure of $N(G) \cap P_n$.

Condition (iii) of Theorem 5.11 implies that given any linear basis $\{g_i\}$ for E_m, i.e. a nonsingular $m \times m$ matrix G, a positive basis may be constructed by adding to the set any vector whose negative lies in int $R_+(G)$. Thus there exist positive bases for E_m of the form

$$\left\{g_1, \ldots, g_m, -\sum_{i=1}^{m} \alpha_i g_i : \alpha_i > 0\right\}, \tag{5.8}$$

with $m + 1$ elements. Since it is easily checked from condition (ii) of Theorem 5.11 that

$$\{G, -G\} \tag{5.9}$$

is a positive basis for E_m, there exist positive bases for E_m with $2m$ elements. It may be shown [16] that the cardinality π of a positive basis for E_m satisfies the inequality

$$m + 1 \leqslant \pi \leqslant 2m. \tag{5.10}$$

A positive basis with $m + 1$ elements is called ***minimal***. It is known that any vector in E_m has a unique representation as a positive combination of the elements of either positive basis (5.8) or (5.9), although this is not true in general [29, 35]. However, it will be seen that (5.9) is the type of positive basis most useful for programming.

It follows from the previous paragraph that all frames for the cone $R_+(G)$ generated by an arbitrary $m \times n$ matrix G must contain a set of at least $l(G) + 1$ vectors in $L(G)$. Selecting such a positive basis for $L(G)$ from the columns of G is complicated, cf. [16]. However, Theorem 5.4 yields a partial solution to the prior problem of finding which columns of G lie in $L(G)$. (Notice that letting $N \doteq N(G)$, we have seen condition (ii) of Theorem 5.11 before as condition (iv) of theorem 5.4.)

THEOREM 5.12. *Let G be an $m \times n$ matrix. Then the following situations are mutually exclusive*:

(i) $N(G) \cap P_n = \{0\}$, *when* $R_+(G)$ *is a pointed cone, i.e.,* $L(G) = \{0\}$,

(ii) $N(G) \cap P_n = R_+(e_1, ..., e_p)$, *when* $L(G) = R_+(g_1, ..., g_p)$ *if* $p \geqslant 2$, *and G contains a zero column if* $p = 1$,

(iii) $N(G)$ *has a basis in int* $R_+(e_1, ..., e_q)$, q *minimal, but* $N(G) \cap P_n \neq R_+(e_1, ..., e_q)$, *when* $L(G) = R_+(g_1, ..., g_q)$ *if* $q \geqslant 2$, *G contains a zero column if* $q = 1$, *and* $g_{q+1}, ..., g_n$ *are a frame for* $R_+(G)/L(G)$,

(iv) $N(G) \cap$ *int* $P_n \neq \emptyset$, *when* $L(G) = R_+(G) = R(G) \subset E_m$.

PROOF: The four situations are mutually exclusive by Theorem 5.4. By Theorem 5.11, if $L(G) \neq 0$, there exists a positive relation between at least $l(G) + 1$ columns of G, which contradicts the assumption of (i). The assertion of (ii) also follows from Theorem 5.11, since $N(G) \cap \text{int } P_p \neq \emptyset$ by assumption. The first assertion of (iii) is proved similarly. Positive dependence among $g_{q+1}, ..., g_n$, or of some of these vectors on $g_1, ..., g_q$ would contradict the hypothesis of (iii) (see (5.6)), which proves the second assertion. The assertion of (iv) is part of Theorem 5.11.

In principle, the columns of G can be divided into three (not necessarily unique) sets: those which form a basis for $L(G)$, those which provide a frame for $R_+(G)/L(G)$ and those which are positively dependent on the others. A general procedure for distinguishing these sets may be based upon Theorem 5.12 and a study of $(I - G^+G)$ [19]. Due to the possible nonuniqueness of the representation of an arbitrary vector in $L(G)$ in terms of its positive basis, however, this partition would appear to be largely of theoretical interest for stochastic programming with linear compensation.

6. STOCHASTIC PROGRAMMING WITH LINEAR COMPENSATION

The remainder of this paper applies the results of Section 5 to a study of stochastic linear programming with linear compensation. In this section necessary and sufficient conditions for the existence of solutions to the program (3.22) are investigated. Without loss of generality, $E\mathbf{c}$ will be replaced simply by c for the remainder of the paper. Many of the results could be extended to the case of convex functionals and first-stage constraints, in particular, to other stochastic criterion functionals (see Section 3).

An examination of the loss (3.21) motivates the following definitions with regard to the set K of points x from which the optimal vector for the program (3.22) is to be chosen. Since any such x must give rise to a corresponding $y_x(A, b) \geqslant 0$ for almost all A, b, a vector $x \in P_{n_1}$ will be said to be *quasi-feasible* (*q-feasible*) if.

$$P(\mathbf{b} - \mathbf{A}x \in R_+(B)) = 1. \tag{6.1}$$

Since to be worthwhile a q-feasible x must give rise to a finite expected loss, a vector $x \in P_{n_1}$ will be called *properly feasible* (p-feasible) if it is q-feasible and the expected value of the corresponding second stage program (2.17) exists.

If $R_+(B)$ is a solid cone, i.e. $R_+(B) = E_m$, the condition (6.1) for q-feasibility becomes trivial, and the set Q of q-feasible x becomes simply P_{n_1}, the positive orthant of E_{n_1}, which is closed and convex. The following theorem shows that these properties are true in general. (Of course, Q may be vacuous.)

THEOREM 6.1. *The set of all q-feasible x, $Q \subset P_{n_1}$, is closed and convex.*

PROOF: Convexity is established by the usual arguments. To prove Q closed, consider a Cauchy sequence $\{x_n\} \subset Q$ with limit x_0. It suffices to show that $x_0 \in Q$. But since P_{n_1} is closed, $x_0 \in P_{n_1}$, and hence it is sufficient to verify (6.1) for x_0. Therefore let $y_n = x_0 - x_n$, so that $\{y_n\}$ is a Cauchy sequence converging to 0, and

$$P(\mathbf{b} - \mathbf{A}x_0 + \mathbf{A}y_n \in R_+(B)) = 1 \qquad \text{for all } n. \tag{6.2}$$

Now suppose $P(\mathbf{b} - \mathbf{A}x_0 \in R_+(B)) < 1$. Then there exists an $\epsilon > 0$ such that $P(d(\mathbf{b} - \mathbf{A}x_0, R_+(B)) > \epsilon) > 0$, where, for A, b fixed, $d(b - Ax_0, R_+(B))$ denotes the minimum distance between the vector $b - Ax_0$ and the set $R_+(B)$ in E_m. Defining $T = \{b - Ax_0 : d(b - Ax_0, R_+(B)) > \epsilon, \ b \in E_m, A \in E_{mn_1}\}$, $P(\mathbf{b} - \mathbf{A}x_0 \in T) > 0$. Now since $y_n \to 0$, there exists an N such that $y_N \in S(\delta)$, an open sphere of radius $\delta < \epsilon$ with centre 0 in E_m. Therefore, if $b - Ax_0 \in T$, $b - Ax_0 + Ay_N \notin R_+(B)$, and hence

$$0 < P(\mathbf{b} - \mathbf{A}x_0 \in T) \leqslant P(\mathbf{b} - \mathbf{A}x_0 + \mathbf{A}y_N \notin R_+(B)),$$

which is a contradiction to (6.2). The theorem follows.

Considerable insight into the nature of the restrictions imposed on $x \in P_{n_1}$ by (6.1), i.e. into the "induced constraints" [15], is gained by an investigation of the situation when A is fixed. Then the following necessary and sufficient condition for q-feasibility is easy to establish using the introductory lemmas of Section 5.

THEOREM 6.2. *When A is a fixed matrix, there exist q-feasible x for the program* (3.22) *iff there exist $z \in R_+(A)$ such that $P(\mathbf{b} \in R_+(B) + z) = 1$.*

In the interests of economy of notation, let

$$\mathcal{Q}_A = \{z : z \in R_+(A), P(\mathbf{b} \in R_+(B) + z) = 1\} \subset E_m , \tag{6.3}$$

and

$$Q_A = A^{-1}[\mathcal{Q}_A] \cap P_{n_1} \subset E_{n_1} . \tag{6.4}$$

By theorem 6.2, there exist q-feasible x in the case under consideration iff $\mathcal{Q}_A \neq \emptyset$. In this case, $x \in E_{n_1}$ is q-feasible iff $x \in Q_A$. The next few lemmas exhibit the structure of $\mathcal{Q}_A$ and Q_A. Without further statement $\mathcal{Q}_A$ will be assumed nonvacuous.

Notice first that when $\mathbf{A}$ is random,

$$P\{(I - BB^+)(\mathbf{b} - \mathbf{A}x) = 0\} = 1, \tag{6.5}$$

for q-feasible x, since (6.1) implies *a fortiori* that $\mathbf{b} - \mathbf{A}x$ lies with probability one in $R(B)$ the orthogonal complement of $N(B')$. When A is fixed, (6.5) may be sharpened by applying $(I - BB^+)$ to any two vectors $b - z_1$, $b - z_2$ (b fixed) generated by a pair of elements of $\mathcal{Q}_A$.

LEMMA 6.3. *All $z \in \mathcal{Q}_A$ have the same component in $N(B')$, i.e., $(I - BB^+) z = \underline{b}^2$, say.*

COROLLARY 6.4. $P((I - BB^+)\, \mathbf{b} = \underline{b}^2) = 1$.

If $R_+(B)$ is solid, these results are trivial, since $N(B') = \{0\}$ and $\underline{b}^2 = 0$. By the methods of Theorem 6.1 one can establish the next lemma.

LEMMA 6.5. *The set $\mathcal{Q}_A \subset E_m$ is closed and convex.*

The special case of Theorem 6.1 when A is fixed follows as a corollary of Lemma 6.5. Indeed, since A is a continuous linear mapping between E_{n_1} and E_m, A^{-1} $[\mathcal{Q}_A]$ is closed and convex. As P_{n_1} is also closed and convex, the special case is established.

The relation (5.1) between $R_+(B)$ and $R_-(B)$ may be used to further describe the nature of $\mathcal{Q}_A$.

LEMMA 6.6. $\mathcal{Q}_A = \{z : z \in R_+(A), P(z \in R_-(B) + \mathbf{b}) = 1$.

Thus $\mathcal{Q}_A$ is generated essentially by the intersection of a fixed closed cone $R_+(A)$ and a randomly translated cone $R_-(B) + \mathbf{b}$. Hence if $\mathcal{Q}_A$ is nonempty, the intersection of $R_+(A)$ and $R_-(B) + b$ for a set of values of $\mathbf{b}$ of probability one must be nonvacuous. If $R_+(B)$ is solid, so is $R_-(B)$, and as before Lemma 6.6 is trivial. Even if $R_+(B) \neq E_m$, $\mathcal{Q}_A$ certainly need not be bounded. Indeed, if there exists a set $\Gamma \subset E_m$ such that $P(\mathbf{b} \in \Gamma) = 1$ and $R_+(A) \subset \cap \{R_-(B) + b : b \in \Gamma\}$, then $\mathcal{Q}_A = R_+(A)$, which is unbounded. In this case, $Q_A = P_{n_1}$ and the set of q-feasible vectors x is also unbounded. More generally, Lemma 6.6 implies that

$$Q_A = \{x : x \in P_{n_1}, P(x \in A^{-1}[R_-(B) + \mathbf{b}]) = 1. \qquad (6.6)$$

When $\mathbf{A}$ is random, whether the set Q of all q-feasible x is vacuous or not depends on the orientation of the random cone $R_+(\mathbf{A})$ and the randomly translated cone $R_-(B) + \mathbf{b}$. The above considerations show that the vector x is q-feasible for the program (3.22) iff one can find a set $\Lambda \subset E_{mn_1}$ such that $P(\mathbf{A} \in \Lambda) = 1$ and $Ax \in R_+(B) + \mathbf{b}|A$ with probability one for all $A \in \Lambda$. Here $b|A$ denotes the r.v. whose distribution is the conditional distribution of the r.v. $\mathbf{b}$ given that the random matrix $\mathbf{A} = A$. (In the present circumstances this distribution always exists, see [20], V. 10, pp. 157-162.) If Λ exists, then Q is essentially given by $\cap \{Q_A : A \in \Lambda\}$. When the distribution of $\mathbf{b}$ is discrete and finite, Lemmas 6.5 and 6.6 and A linear imply that Q_A is a closed convex polytope in E_{n_1}. Thus when the joint distribution of $\mathbf{A}$ and $\mathbf{b}$ is discrete and finite, so is that of $\mathbf{A}$ and $\mathbf{b}|\mathbf{A}$, and Q is a closed convex polytope in E_{n_1}. Alternately, this assertion follows from the fact that in the case under consideration (3.22) is equivalent to a large and complicated linear program [3, 15]. Expression (6.6) shows that, in general, Q is given by

$$Q = \{x : x \in P_{n_1}, P(x \in A^{-1}[R_-(B) + \mathbf{b}]) = 1\}, \qquad (6.7)$$

a closed convex set.

Now if $Q \neq \emptyset$, and the closed convex polytopic cone $R_+(B)$ is not solid, $\mathbf{b} - \mathbf{A}x$ must lie in $R_+(B)$ with probability one for all $x \in Q$. Thus if for each $x \in P_{n_1}$, $\mathbf{b} - \mathbf{A}x$ is distributed over all of E_m, and B generates such a cone, the program (3.22) does not have a q-feasible vector and is thus infeasible. For a B such that $R_+(B) \neq E_m$, the only appropriate distributions of $\mathbf{A}$ and $\mathbf{b}$ are these for which there exists an $x \in P_{n_1}$ such that the distribution of $\mathbf{b} - \mathbf{A}x$ is concentrated in certain half-spaces of E_m. That is, it is necessary (but not sufficient) that the distribution of $\mathbf{b} - \mathbf{A}x$ lie entirely on the cone side of the hyperplanes guaranteed by the (Minkowski-Farkas) Theorem 5.10.2.

We now turn to p-feasibility. It will be shown that when $E\mathbf{A}$ and $E\mathbf{b}$ exist, all vectors which are q-feasible for the program (3.22) are p-feasible under certain necessary conditions on the n_2 vector f and the $m \times n_2$ matrix B of

(3.21). First a sufficient condition for q-feasible x to be p-feasible will be given. Subsequently, it will be shown that depending on the intersections of the complementary pair $\{R(B'), N(B)\}$ with the positive orthant of E_{n_2}, this condition is either trivial or necessary.

Recall that the vector $f \in E_{n_2}$ may be uniquely decomposed into two orthogonal components $f^1 \in R(B')$ and $f^2 \in N(B)$. Now it is a property of the pseudo-inverse of a matrix that all solutions of the constraints of (3.22) are of the form

$$y_x(A, b) = B^+(b - Ax) + (I - B^+B)u, \qquad u \in E_{n_2}, \tag{6.8}$$

for fixed A, b and x. Hence, using the mapping properties of the pseudo-inverse and the duality theory of linear programming, the second-stage program generating the loss (3.21) may be written

$$\begin{aligned} L(b - Ax) - f^1 \cdot B^+(b - Ax) &= \min_u f^2 \cdot u \\ \text{s.t.} \quad (I - B^+B)u + B^+(b - Ax) &\geqslant 0, \end{aligned} \tag{6.9}$$

with dual

$$-\min_w B^+(b - Ax) \cdot w \qquad \text{s.t.} \qquad B^+Bw + f^2 \geqslant 0. \tag{6.10}$$

THEOREM 6.7. *Suppose that* $\mathbf{E}\mathbf{A}$ *and* $\mathbf{E}\mathbf{b}$ *exist and that* x *is* q*-feasible for the program* (3.22). *Then if*

$$f^2 \cdot u \geqslant 0 \qquad \textit{for all} \quad u \in N(B) \cap P_{n_2}, \tag{6.11}$$

x *is* p*-feasible.*

PROOF: By virtue of Lemma 3.1 and the form of (6.9) and (6.10), it suffices to verify (3.5) for the program (6.10), i.e., to show that for almost all A, b, (6.10) is proper. But by the definition of q-feasibility, there exists a feasible vector for (6.9) with probability one. By the duality theory of linear programming, it is therefore sufficient to find a feasible vector for (6.10) for almost all A, b in order to prove the theorem. But by the (Minkowski-Farkas) Theorem 5.10.1 the vector $u \in N(B)$ satisfies $f^2 \cdot u \geqslant 0$ for all $u \in N(B) \cap P_{n_2}$, i.e., for all u such that

$$\begin{bmatrix} B^+B \\ -B^+B \\ I \end{bmatrix} u \geqslant 0,$$

iff there exist $w_1 \in E_{n_2}$ and $w_2 \in P_{n_2}$ such that $f^2 = B^+Bw_1 + w_2$. Hence $w_2 = B^+B(-w_1) + f^2 \geqslant 0$, and taking $w = -w_1$ yields a feasible vector for (6.10). Thus (6.10) and (6.9) are proper with probability one as required.

It will be seen in Section 8 that in certain special cases condition (6.11) may be simplified.

Consider now the four mutually exclusive cases of Theorem 5.4 for the complementary pair $\{R(B'), N(B)\}$ in E_{n_2}. In case (iv), $N(B) \cap P_n = \{0\}$ and condition (6.11) is satisfied by all vectors $f \in E_{n_2}$. Hence

COROLLARY 6.8. *If* $E\mathbf{A}$ *and* $E\mathbf{b}$ *exist and* $N(B) \cap P_n = \{0\}$, *then the sets of p-feasible and q-feasible vectors for the program* (3.22) *coincide.*

For the remaining cases of Theorem 5.4, condition (6.11) is necessary as well as sufficient for q-feasibility to imply p-feasibility.

COROLLARY 6.9. *If* $E\mathbf{A}$ *and* $E\mathbf{b}$ *exist and* $N(B) \cap P_n \neq \{0\}$, *then a vector* x *which is q-feasible for the program* (3.22) *is p-feasible iff* (6.11) *holds.*

PROOF: It remains only to prove necessity. Since x is q-feasible, there exists a feasible vector for the program (6.9) for almost all A, b. Adding any vector $u \in N(B) \cap P_{n_2}$ to this vector yields another feasible vector, so that unless (6.11) holds, (6.9) is improper.

Thus under the condition (6.11), the set of q-feasible vectors Q and the set of p-feasible vectors K coincide. In the remainder of this paper, the term *feasible* will therefore be used to describe both a program of the form (3.22) for which $E\mathbf{A}$ and $E\mathbf{b}$ exist, condition (6.11) holds and $K \neq \emptyset$; and a vector $x \in K$. (Note that a feasible program may still be improper.) In summary, the set K of feasible vectors for the program (3.22) given by (6.7) is closed and convex (as has previously been assumed, e.g., [15]) and, if $\mathbf{A}$ and $\mathbf{b}$ have finite discrete distributions, polytopic. In fact, when A is a fixed matrix, K is in general polytopic, but the proof will be deferred to Section 8.

From a consideration of the proof of Lemma 3.1 and the program (6.10), one sees that $L(b - Ax)$ given by (3.21) is composed of a linear term and a piece-wise linear term given implicitly by the program (6.9) or its dual (6.10). The difficulty in finding solutions to (3.22) (especially when $\mathbf{A}$ is random) lies in obtaining this second term explicitly as a function of the vector x. In Sections 7 and 8, the object is to solve this problem for certain special cases and to study the resulting programs. Generally, however, it is easy to see that $L(b - Ax)$ is a convex (piece-wise linear) function of $x \in K$ for fixed A and b. Hence expected loss is a convex function of x and one has the following lemma [3, 15].

LEMMA 6.10. *The function* $c \cdot x + EL(\mathbf{b} - \mathbf{A}x)$ *is a convex function of* $x \in K$.

It follows that this function is continuous on the relative interior of K as a convex set and upper-semicontinuous at its extreme points. Depending on the nature of the distribution of $\mathbf{A}$ and $\mathbf{b}$, the function may be nonlinear and only piece-wise differentiable.

A criterion for recognizing an optimal $x \in K$ is due to Dantzig and Madansky [15, 32]. It concerns the random vector $p^0(\mathbf{b} - \mathbf{A}x)$ which, upon the realization of $\mathbf{A}$, $\mathbf{b}$, is optimal for the dual of the second-stage program (3.21),

$$\max_p (b - Ax) \cdot p \qquad \text{s.t.} \qquad B'p \leqslant f, \tag{6.12}$$

(alternately, the r.v. $w^0(\mathbf{b} - \mathbf{A}x)$ optimal for the dual program (6.10)). In special cases, it yields the (optimal) value of the program (3.22). These corollaries will be established in the next section. In Section 8, the criterion will be used to obtain optimal vectors for another special case of (3.22). For fixed $x \in K$,

$$c \cdot x + EL(\mathbf{b} - \mathbf{A}x) = [c - E\mathbf{A}'p^0(\mathbf{b} - \mathbf{A}x)] \cdot x + Ep^0(\mathbf{b} - \mathbf{A}x) \cdot \mathbf{b}. \tag{6.13}$$

It is not surprising that the optimality criterion involves only the first term on the right hand side of (6.13), since it is easily demonstrated that

$$[c - E\mathbf{A}'p^0(\mathbf{b} - \mathbf{A}\underline{x})] \cdot x + Ep^0(\mathbf{b} - \mathbf{A}\underline{x}) \cdot \mathbf{b}$$

is a supporting hyperplane to (6.13) at $x = \underline{x}$ [15].

THEOREM 6.11. *Suppose the program* (3.22) *is feasible. Then* $x^0 \in K$ *is optimal iff*

$$c - E\mathbf{A}'p^0(\mathbf{b} - \mathbf{A}x^0) \geqslant 0, \tag{6.14}$$

and

$$[c - E\mathbf{A}'p^0(\mathbf{b} - \mathbf{A}x^0)] \cdot x^0 = 0. \tag{6.15}$$

Thus the optimal choice of the vector $x \in K$ exactly balances $c \cdot x$ against that part of the discrepancy cost under prior control. Conditions (6.14) and (6.15) of Theorem 6.11 are essentially Kuhn-Tucker conditions for the program (3.22). The proof of the theorem uses abstract methods similar to those which Madansky used to prove a slightly different theorem for the case of fixed A [32]. It will be given in another paper. Using the relation

$$p^0(b - Ax) = B'^+f^1 + B'^+w^0(b - Ax), \tag{6.16}$$

for fixed A, b, (cf. (6.9), (6.10), and (6.12)), one obtains the equivalents of (6.14) and (6.15) in terms of $w^0(\mathbf{b} - \mathbf{A}x)$,

$$(c - E\mathbf{A}'B'^+f^1) \geqslant E\mathbf{A}'B'^+w^0(\mathbf{b} - \mathbf{A}x^0), \tag{6.17}$$

and

$$(c - E\mathbf{A}'B'^+f^1) \cdot x^0 = E\mathbf{A}'B'^+w^0(\mathbf{b} - \mathbf{A}x^0) \cdot x^0. \tag{6.18}$$

Finally, (6.9) implies that if the program (3.22) is feasible, it may be written in the form

$$\min_{x \geqslant 0} (c - E\mathbf{A}'B'^{+}f^{1}) \cdot x + E \min_{u} f^{2} \cdot u + f^{1} \cdot B^{+}E\mathbf{b} \tag{6.19}$$

$$\text{s.t.} \qquad P(\mathbf{b} - \mathbf{A}x \in R_{+}(B)) = 1 \tag{6.20}$$

$$(I - B^{+}B)u \geqslant -B^{+}(\mathbf{b} - \mathbf{A}x), \tag{6.21}$$

where for $x \in K$ there exists a feasible u with probability one in (6.21) by virtue of (6.20). In the next section the consequences of the choice of u are irrelevant to the optimal choice of x. The calculations of Section 8 are equivalent to obtaining a unique expression for optimal u in terms of x.

7. Constrained Linear Loss

Let us now investigate the conditions under which the loss (3.21) is linear, but subject to constraints. These constraints serve to limit the domain of the loss function, or alternately, determine the set K of feasible vectors for the program (3.22). Under certain assumptions, decision equivalents of this program and bounds on its value may be obtained. First a general theorem will be given. The proof is an easy consequence of Theorem 6.11 using (6.9), (6.17), (6.18), and (6.19).

Theorem 7.1. *If the program* (3.22) *is feasible and* $f \in R(B')$ (i.e., $f^{2} = 0$), *then expected loss is* $f \cdot B^{+}(E\mathbf{b} - E\mathbf{A}x)$ *for* $x \in K$ *and* (3.22) *is improper unless* $c - E\mathbf{A}'B'^{+}f \geqslant 0$ (cf. Section 2) *when its value is* $f \cdot B^{+}E\mathbf{b}$.

The condition on f of Theorem 7.1 is most reasonable when $R_{+}(B)$ is a pointed cone in E_{m} (i.e., $N(B) \cap P_{n_2} = \{0\}$ or $R(B') \cap \operatorname{int} P_{n_2} \neq \emptyset$, see Theorems 5.4 and 5.12). When $B^{+}B$ is a non-negative matrix as well, Corollary 5.11 may be used to derive decision-equivalents for (3.22). (However, assumptions about the distribution of the random matrix $\mathbf{A}$ will be needed.) When $f \in R(B')$, e.g., when B is of full column rank n_2, these programs become certainty equivalents. More generally, their values yield upper bounds on the value of (3.22).

Suppose that $B^{+}B$ is non-negative. Then by Corollary 5.11, the constraint (6.20) of the program (6.19) may be written as

$$P(\mathbf{b} - \mathbf{A}x \in R(B)) = 1, \tag{7.1}$$

and

$$P(B^{+}(\mathbf{b} - \mathbf{A}x) \geqslant 0) = 1. \tag{7.2}$$

Expression (7.1) is trivial if B is of full row rank m. Expression (7.2) may be interpreted as a chance-constraint in the sense of Charnes and Cooper (cf.

Section 2) which must hold with probability one. Such a constraint is not surprising, for in the case under consideration $R_+(B)$ is not a linear subspace of E_m, i.e., $R_+(B)/L(B)$ is not vacuous. In the remainder of this section $R_+(B)$ will in fact be taken to be a pointed cone, for otherwise certain of the constraints (7.2) (associated with $L(B)$) will be of the form discussed in Section 8.

Let the columns of the $n_2 \times (n_1 + 1)$ random matrix $(B^+\mathbf{A}\ B^+\mathbf{b})$ be independent. Then denoting the marginal distribution function of the i, jth entry of this matrix by F_{ij}, $i = 1,..., n_2$, $j = 1,..., n_1$, and that of the ith co-ordinate of the random n_2 vector $B^+\mathbf{b}$ by G_i, (7.2) may be reduced to a deterministic linear constraint. Indeed, define

$$\gamma_{ij} = -\inf_{\alpha}\{\alpha : F_{ij}(\alpha) = 1\} \qquad i = 1,..., n_2, \quad j = 1,..., n_1 \tag{7.3}$$

and

$$\delta_i = -\sup_{\beta}\{\beta : G_i(\beta) = 0\} \qquad i = 1,..., n_2, \tag{7.4}$$

and consider the ith constraint of (7.2),

$$P\left(\sum_{j=1}^{n_1}\sum_{k=1}^{m} \beta^{ik}\boldsymbol{\alpha}_{kj}\xi_j \leqslant \sum_{k=1}^{m} \beta^{ik}\boldsymbol{\beta}_k\right) = 1, \tag{7.5}$$

where β^{ik} denotes the i, kth entry of B^+. A vector $x \geqslant 0$ satisfies (7.5) iff

$$\sum_{j=1}^{n_1} \gamma_{ij}\xi_j \geqslant \delta_i . \tag{7.6}$$

Hence (7.2) is equivalent to

$$Cx \geqslant d, \tag{7.7}$$

where the entries of the $n_2 \times n_1$ matrix C and the n_2 vector d are given by (7.3) and (7.4), respectively. It follows that for each $x \in P_{n_1}$ which satisfies (7.1) and (7.7), there exists $u \in E_{n_2}$ (e.g., $u = 0$) which satisfies a stronger version of (6.21),

$$(I - B^+B)u \geqslant d - Cx \geqslant B^+(\mathbf{A}x - \mathbf{b}). \tag{7.8}$$

In summary,

THEOREM 7.2. *Let the program* (3.22) *be such that* B^+B *is non-negative,* $R_+(B)$ *is pointed, the columns of* $(B^+\mathbf{A}\ B^+\mathbf{b})$ *are independent,* $E\mathbf{A}$ *and* $E\mathbf{b}$ *exist, and* (7.1) *holds for all* $x \in P_{n_1}$. *Then if there exists a q-feasible vector for* (3.22), C *and* d *exist and* (3.22) *has a decision equivalent of the form*

$$\begin{aligned}\min_{x\geqslant 0}\ & (c - E\mathbf{A}'B'^+f^1)\cdot x + \min_u f^2\cdot u + f^1\cdot B^+E\mathbf{b}\\ \text{s.t.}\ & Cx \geqslant d\\ & Cx + (I - B^+B)u \geqslant d.\end{aligned} \tag{7.9}$$

On the other hand, if C or d does not exist, or (7.9) *is infeasible, then* (3.22) *is infeasible.*

Notice that unless (6.11) is satisfied, (3.22) and (7.9) may have no p-feasible vectors. It follows from the next result that the program (3.22) is proper iff its decision-equivalent (7.9) is proper.

COROLLARY 7.3. *Under the conditions of Theorem* 7.2, *the value of the program* (7.9) *is an upper bound for the value of the program* (3.22).

The proof is immediate from (7.8), which shows that the determination of K is based on extreme values of $\mathbf{A}$ and $\mathbf{b}$ in this case, while the choice of the optimal decision $x \in K$ is based on $E\mathbf{A}$ and $E\mathbf{b}$ and is independent of the consequences of the realization of $\mathbf{A}$ and $\mathbf{b}$. Theorem 6.11 and the convexity of the loss imply that a lower bound for the value of (3.22) may be obtained by the (usual) procedure of solving (6.9) with $A = E\mathbf{A}$ and $\mathbf{b} = E\mathbf{b}$. That is, after finding an optimal x^0 for (7.9), by solving the program

$$\min_u f^2 \cdot u + f^1 \cdot B^+ E\mathbf{b} \qquad \text{s.t.} \qquad (I - B^+B)u \geqslant B^+(E\mathbf{A}x^0 - E\mathbf{b}). \quad (7.10)$$

The first condition of Theorem 7.2 on the distribution of $\mathbf{A}$ is satisfied if the entries of $\mathbf{A}$ are distributed independently of each other and of $\mathbf{b}$, or if B is diagonal and the columns of $(\mathbf{A}\ \mathbf{b})$ are distributed independently. These conditions are somewhat artificial. When A is fixed, the above results can be sharpened using the vector $\underline{b}^2$ of Lemma 6.3 to ensure that the constraint (7.1) is satisfied.

THEOREM 7.4. *Let the program* (3.22) *be such that B^+B is non-negative, $R_+(B)$ is pointed, A is fixed and $E\mathbf{b}$ exists. Then if there exists a q-feasible vector for* (3.22), *d exists and* (3.22) *has a decision-equivalent of the form*

$$\begin{aligned} \min_{x \geqslant 0}\ (c - A'B'^+f^1) \cdot x + \min_u f^2 \cdot u + f^1 \cdot B^+E\mathbf{b} & \\ \text{s.t.} \qquad -B^+Ax \geqslant d & \\ -B^+Ax + (I - B^+B)u \geqslant d & \\ (I - B^+B)Ax = \underline{b}^2, & \qquad (7.11) \end{aligned}$$

whose value is an upper bound for the value of (3.22). *On the other hand, if d does not exist, or* (7.11) *is infeasible,* (3.22) *is infeasible.*

The upper bound of Theorem 7.4 is not necessarily sharper than that given by Madansky [30], p. 200, which is based on the expected loss of taking an optimal compensating decision after basing the first-stage decision on the "expected value" procedure. The present bound is based on an optimal first-stage decision and a "mini-max" compensating decision. It has however the advantage of being more easily calculable (cf. the proof of Lemma 3.1).

Moreover, it is attained in two special cases, the case of B of full column rank n_2, and more generally, if $f^2 = 0$ (when the value of the program (7.11) is $f^1 B^+ E\mathbf{b}$). In the second case, only the first term of any compensating decision (6.8) enters the function to be minimized, although this decision is not known before $\mathbf{b}$ is realized. In both cases, expected loss is a constrained linear function of the first-stage decision x. Similar remarks apply to the case of Theorem 7.2.

When $f \in R(B')$, the above derivations may in principle be applied to the case of general matrices B for which $R_+(B)$ is pointed, i.e., $N(B) \cap P_{n_2} = \{0\}$, using the matrix D given by the (Minkowski-Weyl) Theorem 5.12 in place of B^+ in the derivation of (7.7). In practice however, as mentioned in Section 5, no efficient algorithm for computing D from B exists. Nevertheless, it follows from a consideration of such a matrix for pointed cones, that in determining the set K of feasible vectors for the program (3.22) with A fixed, D would induce only linear constraints, i.e., in this case, K is polytopic. When $f \notin R(B')$, the choice of optimal $x \in K$ is no longer based solely on $E\mathbf{A}$ and $E\mathbf{b}$.

It should be noted that two matrices B which satisfy the above conditions are I_m and $-I_m$. Either of these leads to the fat formulation mentioned in Section 3. When A is fixed and $\mathbf{b}$ has a finite discrete distribution on N points in E_m, Beale, and Dantzig and Madansky have proposed a certainty equivalent for (3.22) in terms of all possible realizations $\{b_i\}$ of $\mathbf{b}$ and their probabilities $\{p_i\}$. This leads, in general, to a large program of the form

$$\min_{x \geqslant 0} c \cdot x + \sum_{i=1}^{N} p_i f \cdot y_i \qquad \text{s.t.} \qquad Ax + By_i = b_i, \quad i = 1,\ldots, N, \tag{7.12}$$

for which decomposition algorithms for the dual problem are useful [15]. (The more general case of (7.10) when the random matrix $\mathbf{A}$ also has a finite discrete distribution appears computationally intractable using these methods.) If B^+B is non-negative and $(I - B^+B)[P_{n_2}] = \{0\}$, the problem (7.12) may be greatly reduced by looking at the images of the vectors b_i under B^+ and applying Theorem 7.4.

8. Piece-Wise Linear Loss

The previous section dealt with the program (3.22) when $R_+(B)$ is a pointed cone in E_m. In particular, B^+B was assumed to be non-negative and the resulting loss was seen to be linear and constrained. In this section, the case at the other extreme, $R_+(B)$ a linear subspace of E_m, will be discussed. It will be assumed throughout that A is a fixed matrix, and, for the first time in this paper, deterministic linear constraints will be explicitly introduced.

When the matrix B has a particular form (which subsumes those treated previously), the resulting program will be reduced to one for which a computational algorithm is known, the loss will be obtained as a piece-wise linear function of $x \geqslant 0$, and in a special case, the optimal decision set out. The paper concludes with some remarks on the nature of (3.22) for general matrices B.

The special case of (3.22) treated in Section 7 was reduced essentially to the "fat" formulation of stochastic linear programming and is meaningful, as we have noted, only when the distribution of the r.v. $\mathbf{b}$ lies within certain bounds. In practice, this means that the results of Section 7 can only be applied to problems in which one type of compensation is represented by the matrix B. For example, if B were positive, the optimal decision x^0 satisfies the "constraints" of (3.4) with probability one, while if B were negative, x^0 causes the "constraints" of (3.4) to be exceeded with probability one. It is true that if B were square and diagonal, both positive and negative compensation could be allowed, but only a preselected one for each constraint. In applications, e.g., in short-range inventory or marketing problems, one often encounters problems in which b is distributed over a linear subspace of E_m, so that B must be such that either type of compensation appears in each stochastic constraint. Thus B must be a positive basis for the appropriate subspace of E_m. The difficulty with an arbitrary positive basis, B, (as noted in Section 5) is that every vector in $R_+(B)$ may not have a unique representation as a positive combination of the columns of B, so that one cannot hope to easily determine the loss (3.21) as a function of x.

Two possible positive bases which give unique representations for any vector in E_m are given by (5.8) and (5.9). There is, however, a practical difficulty with the former. To see this, consider the following matrix B of the form (5.8),

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix},$$

and let $A = I_2$ in the program (3.22). Then the constraints of this simple program become

$$\begin{aligned} \xi_1 + \eta_1 &= \boldsymbol{\beta}_1 + \eta_3 \\ \xi_2 + \eta_2 &= \boldsymbol{\beta}_2 + \eta_3 \end{aligned} \tag{8.1}$$

If $\mathbf{b}$ is, for example, bivariate normally distributed, then assuming x^0 known, it is possible that a $\mathbf{b}$ may be realized that makes $\eta_1^0 = 0$, $\eta_3^0 > 0$ and hence $\eta_2^0 > 0$ in the optimal compensating decision. Thus both types of compensation are induced in the second constraint of (8.1) by the requirements of the first. This is not usually desirable in applications.

In order to be able to appropriately reflect the losses associated with the second-stage compensation, we shall consider a version of the program (3.22) which involves an $m \times 2m$ matrix B of the form $(B_1\ -B_2)$, where B_1 is an $m \times m$ nonsingular matrix, $B_2 = B_1 D$ and $D = \operatorname{diag}(\delta_1, \ldots, \delta_m)$, $\delta_j > 0$. Notice that this matrix is of the form (5.9) since the introduction of D is merely a rescaling of certain coordinates.

Specifically, the program

$$\min_{u \geqslant 0} [c \cdot x + \min_{y_1, y_2 \geqslant 0} E(f_1 \cdot y_1 + f_2 \cdot y_2)] \tag{8.2}$$

$$\text{s.t.} \qquad A_1 x = b_1 \tag{8.2.1}$$

$$A_2 x + B_1 y_1 - B_2 y_2 = \mathbf{b}_2 \tag{8.2.2}$$

will be investigated (cf. [13]). Here x is an n vector, f_i and y_i, $i = 1, 2$, are m vectors, b_1 is an ℓ vector and $\mathbf{b}_2$ is a random m vector. By means of an orthogonal transformation of the stochastic constraints, the more general case when $\mathbf{b}_2$ is distributed over an m-dimensional subspace $R_+(B_1, -B_2)$ of E_k, $m < k$, and B_1 is of rank m may be reduced to the form (8.2.2).

Observe that if $f_1 = 0$, the loss is zero whenever $\mathbf{b}_2$ takes a value b_2 for which $A_2 x^0 \geqslant b_2$. A similar statement can be made for the case $f_2 = 0$, so that this type of program covers the case of zero loss on one side of the constraints. The nonzero vector f_i can be chosen appropriately to reflect the magnitude of the loss on either side of the constraint.

Premultiplying (8.2.2) by B_1^{-1} and absorbing D into y_2, the program (8.2) reduces to the "complete" problem recently studied and reviewed by R. Wets [45],

$$\min_{x \geqslant 0} [c \cdot x + \min_{y_1, y_2 \geqslant 0} E(f_1 \cdot y_1 + D^{-1} f_2 \cdot y_2)] \tag{8.3}$$

$$\text{s.t.} \qquad A_1 x = b_1 \tag{8.3.1}$$

$$B_1^{-1} A_2 x + y_1 - y_2 = B_1^{-1} \mathbf{b}_2 . \tag{8.3.2}$$

It is known that when the distribution of the r.v. in (8.3.2) is either finite discrete, or uniform, the program (8.3) can be treated by existing algorithms. Moreover, in the exponential, or more generally, the absolutely continuous case, the solution may be approximated with known techniques. Wets has utilized methods similar to those of Dantzig and Madansky [15] (for the program (3.22) with A fixed) in order to develop an iterative algorithm for the absolutely continuous case which converges to the solution of the program (8.3). At each iteration, the computations involve the solution of a linear program in $x \geqslant 0$ with constraints (8.3.1). When (8.3) has only stochastic constraints, each such program reduces to an elementary calculation (the minimization of a homogeneous linear form in $x \geqslant 0$) and the computations involved in arriving at a solution to (8.3) are reduced. In the next paragraph, it will therefore be shown how the constraints (8.3.1) and (8.3.2) may be combined in order to reduce the program (8.3) to one which is entirely stochastically constrained.

Note first that the constraints (8.3.1) and (8.3.2) may be written in the form

$$\begin{bmatrix} A_1 \\ B^{-1}A_2 \end{bmatrix} x + \begin{bmatrix} 0 & 0 \\ I_m & -I_m \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ \mathbf{b}_2 \end{bmatrix}, \tag{8.4}$$

and consider a matrix of the form $T = \binom{IV}{0\,I}$. If $\ell \leqslant m$, appropriate zeros may be added to A_1 and b_1, and T may be taken to be $2m \times 2m$ with $V = I_m$. Otherwise, T will be $(\ell + m) \times (\ell + m)$, and V may be taken to be any $\ell \times m$ matrix of rank m such that the $m \times m$ matrix $V'V + I_m$ is nonsingular. In either case, (8.4) holds iff (premultiplying (8.4) by T)

$$\begin{bmatrix} A_1 + VB_1^{-1}A_2 \\ B_1^{-1}A_2 \end{bmatrix} x + \begin{bmatrix} V & -V \\ I_m & -I_m \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} b_1 + V\mathbf{b}_2 \\ \mathbf{b}_2 \end{bmatrix}. \tag{8.5}$$

Indeed, T is nonsingular with inverse $T^{-1} = \binom{I-V}{0\;\;I}$. Now since the matrix $\binom{V}{I_m}$ of the second term is of full column rank, its pseudo-inverse is given by $(V'V + I_m)^{-1}(V'I_m)$, and for fixed x and b_2 the positive coordinates of the unique non-negative solution of the constraints (8.5) are given by the appropriate coordinates of

$$y_1 = Wb_1 + b_2 - (WA_1 + B_1^{-1}A_2)x, \qquad y_2 = -y_1 . \tag{8.6}$$

(Since $(B_1 - B_2)$ is a positive basis for E_m, such a solution exists with probability one for all $x \geqslant 0$, in particular for those which satisfy (8.2.1).) Here the matrix $W = (V'V + I)^{-1} V'$. Hence $y_1 \cdot y_2 = 0$, and denoting by A the $m \times n$ matrix $(WA_1 + B_1^{-1}A_2)$ and by $\mathbf{b}$ the random m vector $Wb_1 + \mathbf{b}_2$, and absorbing D^{-1} into f_2, the program (8.2) has been reduced to a "complete" problem of the form

$$\min_{x \geqslant 0} [c \cdot x + \min_{y_1, y_2 \geqslant 0} E(f_1 \cdot y_1 + f_2 \cdot y_2)] \quad \text{s.t.} \quad Ax + y_1 - y_2 = \mathbf{b}, \tag{8.6}$$

with m stochastic constraints.

Since $B = (I_m - I_m)$ is a positive basis for E_m, all $x \geqslant 0$ are q-feasible for the program (8.6). (Owing to the nature of A when (8.6) has arisen from the program (8.2), nothing is lost by a consideration of vectors $x \geqslant 0$ which do not satisfy (8.2.1).) Let f denote the $2m$ vector $(f_1\, f_2)'$ and consider the feasibility condition (6.11), $f^2 \cdot u \geqslant 0$ for all $u \in N(B) \cap P_{n_2}$, applied to (8.6). For this program, $n_2 = 2m$ and

$$I - B^+B = \frac{1}{2}\begin{bmatrix} I_m & I_m \\ I_m & I_m \end{bmatrix}, \quad \text{so that } f^2 = \frac{1}{2}\begin{bmatrix} f_1 + f_2 \\ f_1 + f_2 \end{bmatrix}.$$

Hence $f^2 \cdot u \geqslant 0$ for all $u \in N(B) \cap P_{2m} = \{\binom{x}{x} : x \in P_m\}$ iff $(f_1 + f_2) \cdot x \geqslant 0$ for all $x \geqslant 0$, i.e.,

$$f_1 + f_2 \geqslant 0. \tag{8.7}$$

The condition (8.7) may also be derived from the dual of the second-stage program of (8.6). Indeed, for fixed x and b, this program becomes

$$\max_p \, (b - Ax) \cdot p \qquad \text{s.t.} \qquad -f_2 \leqslant p \leqslant f_1 , \tag{8.8}$$

which is feasible iff (8.7) holds [45]. It will be assumed in what follows that (8.7) holds and $E\mathbf{b}$ exists, so that the set of feasible vectors for (8.6) is P_n .

Denoting the columns of the $n \times m$ matrix A' by a_i , $i = 1,..., m$, and the ith coordinate of f_2 by φ_{m+i} , it may be seen from an inspection of the program (8.8) that *an* optimal solution $p^0(b - Ax)$ for this program is given by

$$\pi_i^0(b - Ax) = \begin{cases} -\varphi_{m+i} & \text{if} \quad \beta_i - a_i \cdot x \leqslant 0, \qquad i = 1,..., m, \\ \varphi_i & \qquad \text{otherwise,} \end{cases}$$

with corresponding expected value $Ep^0(\mathbf{b} - Ax)$ given by

$$E\boldsymbol{\pi}_i^0 = \varphi_i - (\varphi_i + \varphi_{m+i})F_i(a_i \cdot x) \qquad i = 1,..., m, \tag{8.10}$$

where F_i denotes the marginal distribution function of the ith coordinate of $\mathbf{b}$. It follows that for fixed b the loss generated by (8.8) is a piece-wise linear function of $x \geqslant 0$ given by

$$\max(\beta_i - a_i x, 0)\varphi_i - \min(\beta_i - a_i x, 0)\varphi_{m+i} \qquad i = 1,..., m, \tag{8.11}$$

and expected loss is

$$\sum_{i=1}^{m} \left[\varphi_i E\boldsymbol{\beta}_i - a_i \cdot x E\boldsymbol{\pi}_i^0 - (\varphi_i + \varphi_{m+i}) \int_{-\infty}^{a_i \cdot x} \beta_i \, dF_i(\beta_i)\right]. \tag{8.12}$$

It was noted in section 6 that (8.12) is a convex function of $x \geqslant 0$ whose exact form depends on the distribution of b. The program (8.6) is thus reduced to the convex discrepancy cost program minimizing

$$c \cdot x + EL(\mathbf{b} - Ax) \tag{8.13}$$

over $x \geqslant 0$ with $EL(\mathbf{b} - Ax)$ given by (8.12). The derivations of this paragraph follow Wets, who has studied the form of (8.12) for arbitrary distributions [45]. Expression (8.12) may also be obtained directly from (8.6) [17].

Now consider the necessary and sufficient optimality conditions (6.14) and

(6.15) of Theorem 6.11 for the program (8.6). These become, respectively, $c - A'E\mathbf{p}^0 \geqslant 0$, i.e.,

$$\gamma_j - \sum_{i=1}^{m} \alpha_{ij}[\varphi_i - (\varphi_i + \varphi_{m+i})F_i(a_i \cdot x)] \geqslant 0 \qquad j = 1,..., n, \tag{8.14}$$

and $(c - A'E\mathbf{p}^0) \cdot x = 0$, i.e.,

$$\sum_{j=1}^{n} \left(\varphi_j - \sum_{i=1}^{m} \alpha_{ij}[\varphi_i - (\varphi_i + \varphi_{m+i})F_i(a_i \cdot x)]\right) \xi_j = 0, \tag{8.15}$$

for vectors $x \geqslant 0$ optimal for (8.6). The quantities of (8.14) are the direction parameters of the supporting hyperplane to (8.13) at x (cf. Section 6). If for fixed x and for some j, $1 \leqslant j \leqslant n$, the jth inequality of (8.14) holds, it "pays" to reduce ξ_j, while if its complement holds, it pays to increase ξ_j. A suitable modification of the algorithm of Wets for the program (8.6) is essentially a method of steepest descent applied to the function (8.13) with the variables subject to non-negativity constraints. Notice that when $\varphi_i + \varphi_{m+i} = 0$ in (8.7), $\pi_i^0 = \varphi_i$, which is independent of x, and (from (8.14)) the distribution of β_i plays no part in the determination of the optimal x.

A necessary and sufficient condition for a minimum of (8.13) over *all* of E_n is that the inequalities (8.14) hold with equality. In the absolutely continuous case, these equations are the necessary and (since (8.13) is convex) sufficient differential conditions for an optimum [3, 15, 17]. If the function does not attain its minimum in P_n, at least one of the constraints $x \geqslant 0$ will bind. (Of course, for the program (8.6) to be proper, a minimum of (8.13) need not exist over E_n.) If, on the other hand, the function attains its minimum at $x^0 \in P_n$, then the matrix equation

$$A'(u^0 - f_1) = c, \tag{8.16}$$

must hold, where $u_i^0 = (\varphi_i + \varphi_{m+i})\, F_i(a_i \cdot x^0)$, $i = 1,..., m$. Hence $c \in R(A')$ is a necessary condition for a minimum of (8.13) to be in P_n. Conversely, this condition allows the equation (8.16) to be solved, and in certain circumstances an optimal vector obtained. In practice, the optimal decision will often have strictly positive coordinates. When the program (8.6) has arisen from a problem of the form (8.2) with $m \leqslant \ell, n$, the other conditions required are often met as well. Otherwise, one of the algorithms mentioned above can be applied.

Suppose that $x^0 \in P_n$ yields a minimum of (8.13) and that A is of rank m. Then $c \in R(A')$ and the unique solution of (8.16) is $u^0 - f_1 = A'^{+}c$, i.e.,

$$(\varphi_i + \varphi_{m+i})F_i(a_i \cdot x^0) = \varphi_i - a^i \cdot c \qquad i = 1,..., m, \tag{8.17}$$

where a^i denotes the ith column of the $m \times n$ matrix $A^+ = A'(AA')^{-1}$. Since $0 \leqslant F_i(\beta_i) \leqslant 1$, it follows that

$$-\varphi_{m+i} \leqslant \alpha^i \cdot c \leqslant \varphi_i \qquad i = 1,..., m. \tag{8.18}$$

If $\varphi_i + \varphi_{m+i} = 0$ for some i, $1 \leqslant i \leqslant m$, then equality holds throughout in the appropriate expression of (8.18). Otherwise, if the marginal distribution of β_i is not bounded above, the lower inequality must be strict. A similar remark applies to the upper inequality. In the following it will be assumed that $f_1 + f_2 > 0$ in (8.7). If this is not the case, the operations described below must be performed on those equations of (8.17) for which $\varphi_i + \varphi_{m+i} > 0$, the other equations disregarded, and the corresponding a_i deleted from A. If the F_i, $i = 1,..., m$, are absolutely continuous, the equations (8.17) may be solved uniquely as

$$a_i \cdot x^0 = F_i^{-1}[(\varphi_i - a^i \cdot c)/(\varphi_m + \varphi_{m+i})] \qquad i = 1,..., m. \tag{8.19}$$

Denoting the right-hand side vector of these equations by w^0, $x^0 \geqslant 0$ must satisfy their matrix version,

$$Ax^0 = w^0, \tag{8.20}$$

i.e., $x^0 = A^+w^0 + (I - A^+A)\, z^0$, for some $z^0 \in E_n$. In summary,

THEOREM 8.1. *Suppose the feasible program* (8.6) *involves a* r.v. $\mathbf{b}$ *with an absolutely continuous distribution. If* $x^0 \in P_n$ *minimizes the function* (8.13) *and* A *is of full row rank* m, *then* $c \in R(A')$, *the inequalities* (8.18) *hold, and* x^0 *is a solution of the equation* (8.20). *Conversely, if* $c \in R(A')$, *the inequalities* (8.18) *hold, and there exists a solution* $x^0 \geqslant 0$ *of* (8.20), *then* x^0 *minimizes* (8.13) *and is therefore optimal for* (8.6).

The converse statement is obvious from the above. Notice that there may be many non-negative solutions to (8.20), all differing by a component in $N(A)$, but since $c \in R(A')$, all such components are orthogonal to c and the value of (8.13) is the same for all. Although certain generalizations of Theorem 8.1 are possible, they do not appear to lead to tractable computational algorithms.

It follows from Theorem 8.1 that if the program (8.6) is such that $m = n$, $A = I_m$, $c = 0$ and $f_1 = f_2 \geqslant 0$, the optimal decision x^0 is given by $\xi_i^0 = \max(\underline{m}_i, 0)$, where $\underline{m}_i$ is the median of the marginal distribution of $\boldsymbol{\beta}_i$. Charnes, Cooper and Thompson have given a characterization of a program similar to (8.2) in terms of "generalized constrained hyper-medians" [13].

Finally, in Section 5 it was mentioned that theoretically there exists a three-fold partition of the columns of an arbitrary $m \times n_2$ matrix B into those columns which are a frame for $L(B)$, those which are a frame for

$R_+(B)/L(B)$ and those which are positively dependent on the others. Special cases of the program (3.22) when $R_+(B)$ is a pointed cone were discussed in the previous section. Other special cases when $R_+(B)$ is a linear set have been discussed above. It would appear that in most practical situations involving a program of the form (3.22) for which neither $R_+(B)/L(B)$ or $L(B)$ are trivial, a combination of the models treated here would be appropriate. For such a program, B would be a partitioned matrix and some components of the r.v. $\mathbf{b}$ would be bounded, while others might be unbounded. The analysis of Section 7 could be applied to that submatrix of B spanning $R_+(B)/L(B)$, and the resulting deterministic constraints added to the remaining stochastic constraints to yield a program of the form (8.2). The loss resulting from a feasible decision would thus have both a constrained linear and a piece-wise linear part. A wider class of problems could be treated by considering programs which differ from a problem of this type only by an orthogonal transformation of the constraints, but it appears difficult to conceive a practical situation for which such a model would be appropriate. The device of orthogonal constraint transformation yields, however, the assertion of Section 6 concerning the polytopic nature of the set K of feasible vectors for the program (3.22) with A fixed. Indeed, in Section 7 it was noted that the assertion was valid for pointed cones. Since, by a suitable orthogonal transformation, $L(B)$ and $R_+(B)/L(B)$, a pointed cone, may be placed in different coordinate subspaces of E_m, it follows from the considerations of this section that only $R_+(B)/L(B)$ induces constraints on feasible vectors x, and the general result is established.

9. ACKNOWLEDGMENTS

This paper is based on a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the Carnegie Institute of Technology, Pittsburgh, May, 1965. It is a pleasure to thank Professor M. H. De Groot for his guidance and encouragement through out the preparation of my thesis. The research was supported in part by the Graduate School of Industrial Administration, C.I.T., Resources for the Future, Inc., and, under grants to the Department of Mathematics, C.I.T., by the Ordnance Branch of the U.S. Army and the National Science Foundation of the U.S. The present paper was essentially completed while I was the I.B.M. Research Fellow in the University of Oxford for the year 1965-66.

Note added in proof. The study of stochastic programming with linear compensation has recently been extended to the case of (3.22) with random $\mathbf{f}$ and $\mathbf{B}$ (described as a *stochastic program with recourse*) by D. W. Walkup and R. J. B. Wets, see, e.g., Stochastic programs with recourse. *SIAM J. Appl. Math.* **15** (1967). In the terminology of the present paper, stochastic programs with recourse are discrepancy cost programs with a random loss function. The principal concern so far has been to obtain conditions under which the set of decisions leading to finite expected loss can be captured by inspection of the support set of the joint distribution of the random parameters.

A simplex-based algorithm for making the threefold partition of the columns of a matrix B mentioned in Section 8 has recently been given in R. J. B. Wets and C. Witzgall. Algorithms for frames and lineality spaces of cones. *J. Nat. Bureau Stand.*, Ser. B, **71** (1967), 1-8.

References

1. H. A. Antosiewicz, ed. *Proc. 2nd Symp. Linear Programming.* I and II. Washington, 1955.
2. K. J. Arrow, L. Hurwicz, and H. Uzawa. "Studies in Linear and Non-Linear Programming." Stanford Mathematical Studies in the Social Sciences, Vol. 2, Stanford, 1958.
3. E. M. L. Beale. On minimizing a convex function subject to linear inequalities. *J. Royal Stat. Soc., Ser. B*, **17** (1955), 173-184.
4. A. Ben-Israel. Notes on linear inequalities, I: The intersection of the non-negative orthant with complementary orthogonal subspaces, *J. Math. Anal. Appl.* **10** (1965), 451-493.
5. A. Ben-Israel and A. Charnes. Contributions to the theory of generalized inverses. *J. Soc. Indust. Appl. Math.* 11 (1963), 667-699.
6. F. J. Beutler. The operator theory of the pseudo-inverse. I and II. *J. Math. Anal. Appl.* **10** (1965), 451-495.
7. J. C. G. Boot. The computation of the generalized inverse of singular or rectangular matrices. *Amer. Math. Monthly* **70** (1963), 302-303.
8. A. Charnes and W. W. Cooper. Chance-constrained programming. *Management Sci.* **5** (1959), 73-79.
9. A. Charnes and W. W. Cooper. Chance-constraints and normal deviates. *J. Amer. Stat. Assoc.* **57** (1962), 134-148.
10. A. Charnes and W. W. Cooper. Deterministic equivalents for optimizing and satisficing under chance constraints. *Op. Res.* 11 (1963), 18-39.
11. A. Charnes, W. W. Cooper, and G. H. Symonds. Cost horizons and certainty equivalents: An approach to stochastic programming of heating oil production. *Management Sci.* **4** (1958), 235-263.
12. A. Charnes, W. W. Cooper, and G. L. Thompson. Characterizations by chance-constrained programming. In [24], 113-120.
13. A. Charnes, W. W. Cooper, and G. L. Thompson. Chance constrained studies for linear programming under uncertainty, I. ONR Research Memorandum 120, Carnegie Institute of Technology, Pittsburgh (1964), published in revised form as "Constrained generalized medians and hyper-medians as deterministic equivalents for two-stage linear programming under uncertainty." *Management Sci.* 11 (1965), 83-112.
14. G. B. Dantzig. Linear programming under uncertainty. *Management Sci.* **1** (1955), 197-206.
15. G. B. Dantzig and A. Madansky. On the solution of two-stage linear programs under uncertainty. *In Proc. 4th Berkeley Symp. Math. Stat. Prob.* (I. J. Neyman, ed.) Berkeley, 1961, pp. 165-176.
16. C. Davis. Theory of positive linear dependence. *Amer. J. Math.* **76** (1954), 733-746.
17. M. A. H. Dempster. On stochastic programming. Unpublished doctoral dissertation, Carnegie Institute of Technology, Pittsburgh (1965).
18. M. A. H. Dempster. A near-linear dual for a program of Kataoka. To appear.

19. M. A. H. Dempster. Convex cones and pseudo-inverses. To appear.
20. W. Feller. "An Introduction to Probability Theory and Its Applications, Vol. 2. Wiley, New York, 1966.
21. D. Gale. Convex cones and linear inequalities. *In* [27], Chapter 17, pp. 287-298.
22. D. Gale. "The Theory of Linear Economic Models," New York, 1960.
23. M. Gerstenhaber. Theory of convex polyhedial cones. Chapter 18 in [27], 298-316.
24. R. L. Graves and P. Wolfe. "Recent Advances in Mathematical Programming." McGraw-Hill, New York, 1963.
25. A. L. Householder. "The Theory of Matrices in Numerical Analysis." Blaisdell, New York, 1964.
26. S. Kataoka. A stochastic programming model. *Econometrica* **31** (1963), 181-196.
27. T. C. Koopmans, ed. "Activity Analysis of Production and Allocation. "Cowles Commission Monograph 13. Wiley, New York, 1951.
28. H. W. Kuhn and A. W. Tucker, eds. "Linear Inequalities and Related Systems," Annals of Mathematics Studies, Vol. 38, Princeton, 1956.
29. R. L. McKinney. Positive basis for linear spaces. *Trans. Amer. Math. Soc.* **103** (1962), 131-148.
30. A. Madansky. Inequalities for stochastic linear programming problems. *Management Sci.* **6** (1960), 197-204.
31. A. Madansky. Methods of solution of linear programs under uncertainty. *Opns. Res.* **10** (1962), 463-471.
32. A. Madansky. Dual variables in two-stage linear programming under uncertainty. *J. Math. Anal. and Appl.* **6** (1963), 98-108.
33. A. Madansky. Linear programming under uncertainty. *In* [24], pp. 103-110.
34. P. van Moescke. Stochastic linear programming: A study in resource allocation under risk. *Yale Economic Essays* **5** (1965), 196-254.
35. J. R. Reay. Unique minimal representations with positive bases. *Amer. Math. Monthly* **73** (1966), 253-261.
36. J. K. Sengupta. The stability of truncated solutions of stochastic linear programming. *Econometrica* **34** (1966), 77-104.
37. E. Simons. A note on parametric linear programming. *Management Sci.* **8** (1962), 355-358.
38. H. Theil. Some reflections on static programming under uncertainty. *Weltwirtschaftliches Arch.* **87** (1961), 124-138.
39. H. Theil. "Optimal Decision Rules for Government and Industry." Studies in Mathematical and Managerial Economics I, Amsterdam, 1964.
40. G. Tintner. Stochastic linear programming with applications to agricultural economics. *In* [1], pp. 197-209.
41. G. Tintner. A Note on stochastic linear programming. *Econometrica* **28** (1960), 490-495.
42. A. W. Tucker. Linear inequalities and convex polyhedral sets. *In* [1], pp. 569-602.
43. H. Uzawa. A theorem on convex polyhedral cones. Chapter 2 in [2], pp. 23-31.
44. H. Uzawa. The Kuhn-Tucker theorem in concave programming. Chapter 3 in [2], pp. 32-37.
45. R. Wets. Programming under uncertainty: The complete problem. *Z. Wahrscheinlichkeitstheorie verw. Geb.* **4** (1966), 316-399.
46. P. Wolfe. Methods of non-linear programming. In [24], pp. 67-86.
47. A. Zygmond. "Trigonometrical Series." Dover, New York, 1955.